SCIENCE LIBRARY
GENERAL LIBRARY
APR-91921
UNIV. OF MICH.

Vol. III, No. 6 PSYCHOLOGICAL REVIEW PUBLICATIONS December, 1920

Journal of Experimental Psychology

EDITED BY

JOHN B. WATSON, New York City
HOWARD C. WARREN, Princeton University (*Review*)
JAMES R. ANGELL, New York City (*Monographs*)
SHEPHERD I. FRANZ, Govt. Hosp. for Insane (*Bulletin*) and
MADISON BENTLEY, University of Illinois (*Index*)

ADVISORY EDITORS

H. A. CARR, University of Chicago; RAYMOND DODGE, Wesleyan University; KNIGHT DUNLAP, The Johns Hopkins University; H. S. LANGFELD, Harvard University; E. L. THORNDIKE, Columbia University.

CONTENTS

Psychological Tests for Selecting Aviators: George M. Stratton, Henry C. McComas, John E. Coover, and English Bagby, 405.

The Relative Value of Grouped and Interspersed Recitations: E. B. Skaggs, 424.

Results of Some Experiments on Affection, Distribution of Associations, and Recall: C. H. Griffitts, 447.

Two New Time Control Instruments: Donald A. Laird, 465.

PUBLISHED BI-MONTHLY BY
PSYCHOLOGICAL REVIEW COMPANY
41 NORTH QUEEN ST., LANCASTER, PA., AND
PRINCETON, N. J.

Application for entry as second-class matter at the post-office at Lancaster, Pa., pending

Psychological Review Publications

EDITED BY

HOWARD C. WARREN, Princeton University (*Review*)
JOHN B. WATSON, 244 Madison Ave., New York (*J. of Exp. Psych.*)
JAMES R. ANGELL, 522 5th Ave., New York (*Monographs*)
SHEPHERD I. FRANZ, Govt. Hosp. for Insane (*Bulletin*)
MADISON BENTLEY, University of Illinois (*Index*)

WITH THE CO-OPERATION OF
MANY DISTINGUISHED PSCHOLOGISTS

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bimonthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 480 pages.

PSYCHOLOGICAL BULLETIN

containing abstracts, critical Reviews, of books and articles, psychological news and notes, university notices, and announcements, appears monthly, the annual volume comprising about 720 pages. Special issues of the Bulletin consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bimonthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 480 pages.

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The Index is issued annually in May, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL MONOGRAPHS

consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The Monographs appear at irregular intervals and are gathered into volumes of about 500 pages.

Philosophical Monographs: a series of treaties more philosophical in character.

Library of Genetic Science and Philosophy: a series of bound volumes.

ANNUAL SUBSCRIPTION RATES

Review: $4.25 (Foreign, $4.50). **Review and Bulletin:** $9.00 (Foreign, $9.50).
Journal: $4.25 (Foreign, $4.50). **Review and Journal:** $8.25 (Foreign, $8.75).
Bulletin: $5.00 (Foreign, $5.25). **Journal and Bulletin:** $9.00 (Foreign, $9.50).
Any one of above with Index: $1.25 additional.
Any two of above with Index: $1.00 additional.
Review, Bulletin, and Journal: $12.50 (Foreign, $13.25).
Review, Bulletin, Journal, and Index: $13.50 (Foreign, $14.25).
Current Numbers: Review or Journal, 80c; Bulletin, 50c; Index, $1.50.
Psychological Monographs: $5.30 per volume (Foreign, $5.80).
Current Issues: prices vary according to size.

Subscriptions, orders, and business communications may be sent direct to the

PRINCETON, N. J. OFFICE

PSYCHOLOGICAL REVIEW COMPANY

FOREIGN AGENTS: G. E. STECHERT & CO., London (2 Star Yard, Cary St., W.C.)
Paris (16, rue de Conde)

Journal of
Experimental Psychology

VOL. III, No. 6. DECEMBER, 1920

PSYCHOLOGICAL TESTS FOR SELECTING AVIATORS

BY

GEORGE M. STRATTON, Major, Air Service (Aeronautics)
HENRY C. McCOMAS, Captain, Sanitary Corps
JOHN E. COOVER, Captain, Sanitary Corps, and
ENGLISH BAGBY, 1st Lieutenant, Sanitary Corps

Medical Research Laboratory, Hazelhurst Field, Mineola, Long Island, New York

PART I. INTRODUCTION

The research here reported[1] aimed to add to the means already available of eliminating very early those candidates for training in aviation who would later be unable to learn to fly, and of selecting those who possessed unusual aptitude for flying. The research did not aim to assist in the classification of those who were already in the service and who had in practice made clear that they were competent flyers.

An understanding of the conditions under which the experiment was undertaken will be assisted by a brief statement of:

[1] Responsibility for this investigation and its report is as follows: For the general direction of the research, for part of the apparatus and experimental methods, and for writing Part I., Major Stratton; for part of the apparatus and methods and for the administration of the experiments at the aviation fields, for writing the description of methods, a summary of results, and the method of determining the aviation rating, Captain McComas and Lieutenant Bagby; for the entire statistical treatment, for writing the section on the judgment of curves, the detailed results, and suggestion on method, and for assistance in revision, Captain Coover.

Acknowledgment should be made to Captain H. M. Johnson for assistance in preparing the report; to Captain M. Bentley for the form of the Maze Experiment; and to the Commanding Officers, of Taylor Field, Montgomery, Ala., of Souther Field, Americus, Ga. (where the experiments were conducted), and of the Medical Research Laboratory at Mineola, for their hearty coöperation.

1. The means earlier used for the selection of men for training in aviation.
2. An enumeration of the psychophysical tests out of which the present study grew.
3. A table of the psychophysical qualities assumed to be important for aviation, upon which the present research was based.

1. The means earlier used included (*a*) the standard physical and medical examination administered by the Aviation Examining Boards and known as 'Examination 609.' This covered acuity of vision, color vision, the balance of the ocular muscles, the normality of the semicircular canals, normal acuity of hearing, and a large number of other matters. (*b*) There was the professional and mental examination administered by the Aviation Examining Boards, based upon the candidate's written answers to a series of questions covering his parentage, education, business experience, athletic attainments, responsibilities placed upon him by others, military training,—his answers being accompanied by letters from at least three persons who could speak with particular knowledge of his moral character and of his fitness to enter the Air Service. The candidate was also required, as a part of this professional and mental examination, to furnish credentials of his education and to appear for oral examination by the Board.

2. As a result of an earlier study,[1] certain tests had been found to be of aid in detecting ability to fly. The tests which proved of value were those on the perception of gradual tilt; on the power to stand steadily, as judged by the record which a man makes when a writing point attached to his head moves over a smoked surface; on his power quickly to discriminate between a sudden jerk of his body to the right or to the left, particularly when this is combined with his reaction time to a visual signal and to an auditory signal; and on the steadiness of his hand when a pistol shot is fired behind his back.

[1] See Henmon, V. A. C.: 'Air Service Tests of Aptitude for Flying,' *J. of Appl. Psychol.*, 1919, 3, 103. Thorndike, E. L.: 'Scientific Personnel Work in the Army,' *Science*, 1919, 49, 53. Stratton, G. M.: 'Psycho-physical Tests of Aviators,' *The Scientific Monthly*, 1919, 8, p. 421.

The tests which did not scientifically justify themselves (at least with the particular apparatus and methods then employed) were those upon a person's power to learn certain complicated combinations of movement of hand and foot, on the power to continue in imagination a fragment of a curve presented to him in model, and on dexterity. This latter test was disapproved not so much because it arrived at nothing which could be connected statistically with flying ability, as that it could so largely be influenced by practice, and practice would be invited if the test were introduced as a regular part of a Board's examination, when the candidates would soon know beforehand that they would be tested on this feat of dexterity.

The earlier research therefore had its measure of success. Tests were actually found by which it was possible to give a more assured judgment than was possible by credentials and personal interview alone, that the candidate could learn to use with skill the sound and normal body which the medical examination gave assurance that he possessed. These earlier experiments also brought home to one the exceeding complexity of the flyer's ability. His power to attain success in his hazardous and delicate work evidently is not due to any single factor in his psychophysical constitution, nor to any small group of factors, but rather to a happy constellation of many factors, of which each counts only in a limited way.

Upon the basis largely of this earlier research certain tests were approved by the Director of Military Aeronautics, and were concerned with the following: Simple reaction time to visual signal, simple reaction time to auditory signal, discriminatory reaction time to sudden tilt, steadiness of standing, emotional stability under surprise, sensitivity to very gradual tilting of the body, and mental alertness. There was also required with these special tests a detailed report of the amount and direction of the candidate's athletic performance, supported, if need be, by the testimony of others.

3. In continuing the research already mentioned, no attempt was made to attack the particular problems connected with the various specialties of aviation, such as bombing,

pursuit, and artillery control. All of these presuppose ability to fly, and the present research was confined to this ability. In planning the experiments the following analysis served as a guide in the selection of the tests to be studied. With the several qualities here assumed to be of importance, are stated the special—and, one must acknowledge, very insufficient—means for determining whether they were present or absent in a given individual. The new tests studied in the present investigation (indicated by their appearing in italics) will thus be seen in their connection with the tests already officially approved and adopted, as well as with the other means of examination.

Qualities Entering into the Ability to Fly	*Means of Their Detection*
1. Suitable Volitional and Emotional Organization.	Standard mental and professional Examination: "Examination 609."
Strength, constancy, and proper direction of purpose; readiness to meet danger; coolness in emergency; zest for aviation in particular.	Psychiatric examination. *Test of muscular exertion and endurance.* Test of emotional stability; *Test of the rapidity of complex reactions.*
2. General Intelligence and Training.	Standard mental and professional examination; Progress in ground school; Special tests of mental alertness.
Quickness, discipline of judgment information.	
3. Special Intellectual Equipment.	*Test of judgment of curves.*
Sufficient span and steadiness of attention; power to discriminate and to remember the parts of the airplane; spatial systematization.	*Test of learning and recall of pathways.* *Test of judgment of relative speed.*
4. Sensory and Perceptive Equipment.	
Normal vision with respect to:	
acuity	
'depth' perception	
binocular	'609.'
monocular	
sensitivity to motion in the visual field	
twilight vision	
color vision.	'609.'
Normal hearing with respect to:	
perception, recognition, and discrimination of	
intensities of sound	'609.'
qualities of sound	
localization of sound	
perception of motion by the internal ear.	'609.'

Normal tactile and 'organic' sensitivity.	'609.' Test of discriminatory reaction time to sudden tilt; Test of sensitivity to very gradual tilting of the body; Test of steadiness of standing.
5. Psycho-motor Equipment. Strength, endurance, quickness; skill in learning precise coördination.	'609.' *Test of muscular exertion and endurance.* Test of simple auditory and visual reaction-time. *Test of rapidity of complex reactions.*

Part II. Methods and Results

1. *Judgment of Curves*

The ability of the aviator to estimate the course that his airplane will take as it descends is of great importance, for bad judgment in landing is responsible for many crashes. This ability undoubtedly depends in part upon acuity of vision, balance of the ocular muscles, judgment of spatial relations, all of which are engaged in the work of this test, namely: that of estimating the distance of the point at which a parabolic curve,[1] if continued beyond a given segment, intersects a horizontal plane. The segment of the curve is presented on the upper edge of a metal plate $\frac{1}{8}$ inch in thickness, which is set vertically in the apparatus, so that the OX-axis (at which $y = 0$) lies in the plane of the upper surface of the base of the apparatus. A millimeter scale upon this surface enables the experimenter to read the position of the pointer with which the aviator indicates his estimate. The indicator is a pointer clamped to an endless cord, held taut between pulleys, and is freely manipulated by a knob.

The apparatus is $56\frac{1}{4}$ inches (142.9 cm.) long; the plates are 15 inches long and 15 inches or less in height. The range of excursion of the indicator is from 0, at the end of the plate nearer the observer, to 39 inches (99 cm.) within $2\frac{1}{4}$ inches of the farther end of the base. When the plate is in position, the OY-axis (at which $x = 0$) lies 14 inches back of the origin of the millimeter scale and within $1\frac{1}{3}$ inches of the back end of the base. The headpiece permits a view with the eyes about $23\frac{1}{4}$ inches above the base of the apparatus, or $8\frac{1}{4}$ inches above the highest plates, and approximately in the plane of the OY-axis.

[1] It is of course not to be understood that an airplane in landing follows such a curve.

The general formula for the parabolic curves is $(y - a)^2 = 4p(x - b)$. The following table gives the constants for the equations from which the six curves were derived:

TABLE I

CONSTANTS IN EQUATIONS FOR PARABOLIC CURVES

Curves	1	2	3	4	5	6
a	18.06	20	7.82	12.94	14.94	30
b	5	0	0	0	−3	0
p	2.087	4	0.46	1.92	1.704	11.25
i	44.0	25.0	33.25	21.75	29.76	20.0
i_1	111.8	63.5	84.5	55.2	75.5	50.8
i_2	76.3	28.0	49.0	19.7	40.0	15.3

The values in the table above the line are given in inches; in the lower part, in centimeters. Since the estimates of the point of intersection (i) were read from a millimeter scale originating at the front of the plate which lies 14 inches, or 35.5 cm., from the origin of the curve (where $x = 0$), the distances from the origin of the curves to the points of intersection (i) had to be converted into values comparable to the estimates; these are given in the last line of the table (i_2).

The administration of the test provided for one sitting on each of two different days. At each sitting the aviator was given the plates first in 1–2–3 order and then in an amended reverse order; consequently two estimates were made on each plate in a sitting, and they were not consecutive. In making the estimate the aviator, with his eyes approximately in the plane of the OY-axis, glanced down, with binocular vision, upon the curved edge of the plate, and out along the base, and manipulated the indicator until it marked the point at which he judged that the curve if continued would meet the plane of the base. The eye-shield, which resembled that of the common stereoscope, confined the shifting of the point of regard within narrow limits, and the screen into which the eye-shield was set prevented any other view of the curves as related to the base. Consequently the criteria on the judgment were limited to the binocular view of the segment of the parabolic curve on the plate, and of the base, from this point of regard.

Results.—The estimates on each of the parabolic curves were tabulated in frequency-distributions, and two facts at once appeared:

1. There was a general tendency to overestimate the distance; and

2. Dispersion in all of the distributions is large.

The mean amount of overestimation, if we neglect Curve 1, ranges from 25.2 per cent. to 87.5 per cent. of the distance estimated.

The amount of dispersion may be shown by (*a*) the range of estimates, and (*b*) the standard deviation (σ) in Table II.

TABLE II

DISPERSION

(*Values in Centimeters*)

Curve	*i*	Range		σ		*P E* of Est.	
		Taylor	Souther	Taylor	Souther	Taylor	Souther
1.....	76.3 cm.	60	80	12.5	15.5	8.4	10.5
2.....	28.0	65	60	11.4	10.9	7.7	7.4
3.....	49.0	80	90	15.9	17.7	10.7	12.0
4.....	19.7	65	55	10.8	9.3	7.3	6.3
5.....	40.0	70	70	16.2	13.7	10.9	9.3
6.....	15.3	45		8.4		5.7	

For the sake of comparison the distance estimated (*i*) is also shown. The range of estimates, with but one exception, is greater than the distance estimated; and on Curve 4 is three times as great. The standard deviation (σ) is from 1/5 to 1/2 of the distance estimated. Consequently the probable error of an estimate (*PE* of est.) is large: for Curve 1, for example, the least distinguishable distance (a value equal to twice the *PE* of est.) is about 18 cm., or a quarter of the distance estimated; and for Curve 4 it is about 14 cm., or a third of the distance estimated.

The general overestimation indicates the presence of a normal illusion of spatial judgment and suggests the necessity for scoring the observers in terms of *deviation from the mean estimate*, rather than from the point of intersection. And the large probable error of the individual estimates suggests that

the correlation of any values based upon them, and aviation rating, would be low.

Although the dispersion of estimates on a curve is large, the four estimates of a single observer commonly vary within a narrower range. For Souther Field, for example, there is a high correlation between the first and second estimates on Curve 1 ($r = .710 \pm .028$) and a fair correlation between the estimates of the two days ($r = .662 \pm .030$). *Consistency of judgment* was rated by MV/M, in which MV is the observer's mean variation and M his own mean.

The coefficients of correlation between aviation rating and the deviation from the mean estimate are shown for the six curves separately in lines 1 to 6 of Table III.: between aviation rating and deviation from the mean divided by the standard deviation (x/σ) for all curves, in line 7; between aviation rating and the mean variation divided by the Mean (MV/M), in line 8; between aviation rating and deviation from the distance to the point of intersection ($E - i$), in line 9.

TABLE III

CORRELATIONS OF JUDGMENTS OF PARABOLIC CURVES, WITH AVIATION RATING

Line.	Curve	Taylor Field	Souther Field
1.....	1	.119 ± .047	.135 ± .079
2.....	2	.142 ± .046	−.025 ± .080
3.....	3	.120 ± .047	.192 ± .077
4.....	4	.149 ± .046	.204 ± .080
5.....	5	−.056 ± .047	.195 ± .080
6.....	6	.140 ± .047	
7.....	Dev. from Mean Estimate (x/σ)	.109 ± .038	.253 ± .035
8.....	Consistency of Judgment (MV/M)	.139 ± .038	−.013 ± .036
9.....	Dev. from Point of Inter. ($E - i$)	.037 ± .020	.052 ± .035

The results with the two sets of observers are in fair agreement: for the single curves, they agree in rank-order of the coefficients for Curves 4, 3, and 1, while the significance of their disparity for Curves 2 and 5 is largely removed by the appearance of two negative but unreliable coefficients among the set of only moderately reliable, and low, coefficients. Both show a reliable but low correlation between aviation rating and the deviations from the mean estimate when the values from all curves are aggregated (line 7), and show the

absence of a reliable correlation with the deviations from the distance to the point of intersection (line 9). A reliable but low correlation between aviation rating and 'consistency of judgment' is found for the observers at Taylor Field, but none for the observers at Souther Field.

On the whole, those aviators whose estimates more closely approximate the 'normal' estimate, and those who are more consistent in their estimates, stand a slightly better chance of a favorable rating in aviation than do their fellows. Since the cadets are a selected, fairly homogeneous class, the low but reliable coefficients of correlation presented above may be taken to indicate that the special abilities tested contribute in some measure to aviational ability. That the correlations with aviation rating are not higher, may depend upon certain disabilities, applicable to the other tests as well, which will be noted in the closing paragraphs of Part III.

2. *Judgment of Relative Speeds*

The aviator is frequently required to estimate the speed of one ship with reference to the speed of another. He must often estimate the speed of his own ship relative to stationary objects also. It is important to learn whether visual judgments of this nature are as greatly influenced by the personal equation as are other types of visual judgment. If large differences in this particular exist, it is important to discover their relation to flying ability in general.

The question was tested by placing before the cadet two moving objects and determining his estimates of their relative speeds.

The apparatus employed may be briefly described as follows.

Two white spots, emerging from behind a screen, travelled along two straight converging lines in a horizontal plane, upon which the observer, standing, looked down. These white spots in due time disappeared behind a second screen which concealed also the common point (*d*) toward which they converged.

Each cadet was given two sittings, each of which consisted

of two series, of 10 presentations each. At each presentation the subject was required to judge whether the white spot on the right would pass *d* before, after, or simultaneously with the white spot on his left. The moving spots were visible during their passage through a distance of 55 cm. between the two screens.

In Series *A*, the spot on the subject's left moved at a rate of 3.58 cm. per second, and the one on the right at the rate of 4.46 cm. per second. In Series *B*, the spot on the right travelled 5.34 cm. per second while the one on the left traveled at 3.58 cm. per second. The spots were arranged to pass the point *d* in varying relations to each other: the right-hand spot was set to pass ahead of, or after, or to 'collide' with, the left.

Results.—Errors were made most frequently when the spots were set to pass a short distance apart, or to 'collide.' There are few errors in the ± 5 cm. settings, but there are many in the ± 2.5, the ± 1.6, and the 0 ('collide') settings.

A comparison of the errors for the different settings of the different days and series shows a typical and persistent error. If the errors are arranged on the basis of overestimation and underestimation of the speed of the right-hand spot with reference to that of the left, we get the following table for the two fields.

TABLE IV

NUMBER OF ERRORS IN JUDGMENT OF SPEED OF RIGHT-HAND SPOT

Taylor Field

	Series A		Series B	
	Day 1	Day 2	Day 1	Day 2
Overestimated	21	16	14	9
Underestimated	69	42	99	103
Souther Field				
Overestimated	32	42	30	14
Underestimated	69	107	53	143

It appears that there is a strong tendency to underestimate the speed of the right-hand spot, which always moves faster than the left. Possibly this may be correctly described as an overestimation of the speed of the slower with respect to the

faster. This tendency increases when the right spot increases its speed, as in Series *B*. The error is persistent, and occurs more frequently in the last sitting of the test than in the first.

It would be interesting to follow this type of illusion in experiments of a different character, and to determine how consistent it is under varying conditions. If the tendency is general, as the uniformity of the results here indicate, the fact is important, as the effect is directly contrary to other effects of 'contrast,' and it should certainly be taken into account in aviational training.

The presence of this illusion complicated the scoring and treatment of the results. If the errors caused by it are frequent, relatively to the errors caused by bad judgment, they, if given the same weight, would tend to obscure the judgmental differences which should be the exclusive basis for scoring.

The likelihood of making any particular error, with each setting of the apparatus, was determined by a tabulation of all of the errors, and each cadet was penalized for an error by a score numerically equal to the number of his fellows who did not make it. Again the graver of two errors, possible in some settings of the apparatus, was more heavily weighted: thus the test-score was freed from the influence of the illusion; and, on the basis of frequency and degree of error, it fairly placed the cadet with respect to normal judgment of speeds.

When the scores are compared with the flying ratings it appears that the tests at both fields yield a positive and moderately reliable correlation; at Taylor Field, $r = .232 \pm .096$; at Souther Field, $r = .224 \pm .085$. These figures to some degree indicate that the better flyer possesses an ability to estimate the relative speed of two moving objects with greater accuracy than does the poorer flyer.

3. *Complex Reaction Time*

The following test was designed at the suggestion of several practical flyers. These men did not agree in their opinions concerning all of the traits the good aviator should possess, but they did agree that he must be quick and accurate

in his control of his machine. It seemed advisable, therefore, to test the cadets in reference to the speed and correctness of the movements which they make in obedience to certain signals.

This was done by the determination of a form of 'discrimination time' which was registered by movements somewhat similar to those that the aviator makes in controlling his plane.

The apparatus consisted of a seat, stick, and rudder-bar, taken from a dismantled airplane and assembled in the same relations that obtained in the airplane. The stick and rudder were equipped with electrical contacts which were wired in parallel with each other and in series, with a dry-cell battery, and the magnet actuating the marker of a Foote-Pierson Writing Register. The subject sat before a screen resting on a table. Behind the screen was a shutter, which was thrown open by the operator so as to expose a card through the opening in the screen. The cards were 14 cm. x 11 cm. On each was a letter and an arrow. The direction of the arrow indicated the foot to be thrust forward. The letter indicated the direction the stick should be moved, thus *R*, to the right, *L*, to the left, *B*, backward and *F*, forward.

The shutter was equipped with an electric contact which closed the circuit through the stick and rudder the instant the shutter was opened. This circuit was broken when both the stick and the rudder moved. The current, therefore, ran through the marker from the time of the opening of the shutter until both the stick and rudder were moved. The marker was actuated by a magnet and traced an ink line, while another marker, controlled by a Jaquet clock, registered time in units of 0.2 seconds. To determine the time between the appearance of the signal and the reactions of the cadet, the length of the continuous line was compared with the number of time-divisions registered beside it. The length of the time-marks was great enough to permit of reading the times in units of 0.1 seconds.

At Taylor Field each cadet was given a number of trials to accustom him to the apparatus before the tests began.

When he had learned the procedure, the time was taken for each reaction, and the correctness of his movements on stick and rudder noted. He was required to make 20 reactions at each sitting on two successive days.

The time of each reaction was measured and an average time of reaction for the two series was taken for each cadet; also the number of incorrect reactions. A comparison of these reaction-times and errors shows that the quickest man made his reaction in 0.64 seconds, and the slowest in 1.64 seconds; that the most accurate made no errors on either day, and the least accurate made sixteen errors the first day and thirteen the second, out of the twenty tests for the day. For the correct reactions, in series which contained not more than five per cent. of wrong reactions, the average time was 1.1 second.

The first and most natural comparison is that between the average time of correct reaction, and the percentage of errors in reacting. This comparison shows that the men who make the greatest number of errors have the shortest reaction time ($r = -.415 \pm .06$ for Taylor Field, and $r = -.337 \pm .07$ for Souther Field). The relation between aviation rating and the average time of the correct reactions ($r = .157 \pm .07$) is partially masked by the influence of the errors. By imposing a penalty of 2.3 per cent. in time for 1 per cent. of error in reaction (coefficient of regression of time on error in reaction = .023) we find the relation between the revised scores and aviation rating to be somewhat higher ($r = .170 \pm .067$) and by excluding the disturbing influence of the errors by a more adequate method of 'partial correlation' we find a reliable correlation ($r = .256 \pm .064$).

At Souther Field a different procedure was followed in preparing the cadets for the tests. They were given a preparatory series of but three trials before the actual testing began. In many cases, therefore, the men were not fully accustomed to the test before the times of reaction were taken. Their results show no correlation with aviation rating.

4. *Tests of Learning and Recall of Complicated Pathways*

Individuals differ widely in the ability to remember a route once traversed, when required to retrace their course. As this trait is often serviceable in cross-country flying, a series of measurements was attempted, by the use of a maze to detect variations in the ability and to determine its correlation with aviation rating.

The apparatus consisted of a maze traced on a cardboard 16 x 22 cm., the paths being 1.5 cm. wide. Over this cardboard rested a covering board with a circular opening 2 cm. in diameter. For preliminary practice the cadet was presented with a smaller and far simpler maze, the entrance to which was indicated, and he was instructed to follow the winding paths from the entrance to the goal, and return. The complete course was to be covered without error in the shortest possible time. The larger maze was then presented to the cadet, who took the covering board, looked through the opening at the entrance and sought to grope his way to the goal by moving the opening over the paths. If he chose a blind alley he returned to the entrance of the alley. He was forbidden to cross a line.

The time was taken in units of 0.2 seconds for the route in to the goal, and again for the return. The test was repeated once on the same day and twice on the following day. The times for ingress and egress for the several trials were compared. In general the memory of the route found in the first tests shortened the times of the later tests. No reliable correlation was found between this and flying ability. Among the possible explanations an obvious one is that the cadets were not ranked for cross-country flying; and, even if the test were of real value in detecting a retentive memory for routes covered, it was not compared with flyers ranked on the basis of such an ability.

5. *Tests of Muscular Exertion and Endurance*

In many situations the aviator is called upon to exercise considerable physical strength and endurance. It was not

practicable to test the cadets for strength of all groups of muscles. As the strength of grip is frequently taken by directors of physical culture as a general index of strength, it was deemed advisable to use the grip of hands in a series of tests.

The apparatus used was the Smedley hand dynamometer from which the dial and hands had been removed, and a rotating drum with a pencil marker substituted. The strength of grip was indicated by the pencilled tracing on a ribbon of paper wound upon the drum. The height of the line indicated the strength of the grip. If the cadet held his grip while the drum revolved, the marker traced a temporal record of the relative degree of maintenance of grip.

The maximal momentary grip was taken separately for the right and left hands. The cadet was then directed to grasp the instrument with his right hand, as tightly as he could, and to maintain the grip for three minutes. The test was repeated the next day with the same procedure.

A comparison of these records shows a wide difference in physical strength. At Taylor Field the strongest right-hand grip was 83 kilograms; the weakest 44 kilograms; the strongest left-hand grip was 74 kilograms; and the weakest 37. Several of the men showed equal strength of grip with the two hands.

The curves for the sustained grip were measured by taking the height of the curve at the beginning of the test and after the lapse of each 22.5 seconds thereafter. This gives five measurements within the first $1\frac{1}{2}$ minutes, after which the curves are too erratic for useful comparison.

These five measures show a consistency in decline for men of different initial strengths, the stronger men losing relatively more strength than the weaker men.

There is no significant correlation between the maximum grip of the right-hand and aviation grade ($r = -.034 \pm .09$ for Taylor Field and $r = -.084 \pm .09$ for Souther Field). Nor is there any correlation between aviation ability and the summated strength grips of the two hands; or the difference between right and left, or the endurance shown in sustained

grips ($r = -.161 \pm .09$ Taylor Field, and $r = .148 \pm .098$ Souther Field).

It would appear that differences in strength (of the forearm, at least) do not differentiate aviators whose flying abilities were as similar as those of the 48 who took this test at Taylor Field and the 55 at Souther Field.

Part III. Correlation between Test-scores and Aviation Rating

1. *Aviation Rating*

At Taylor Field the instructor of each cadet was requested to give separate estimates on each of the following features of flying:

(*a*) Dual Flights (each day)
(*b*) " " (" week)
(*c*) Solo " (" day)
(*d*) " " (" week)

The following grades were used for the purpose: Fair, Good, Very Good, and Excellent.

From the estimates secured from the instructors, the cadets were divided into eight groups for the purposes of correlation. These are designated by the letters *A* to *H*, *A* being the group of greatest efficiency. In the following table the number of cadets falling into each group is indicated.

Group	No. of Cadets
A	1
B	5
C	7
D	14
E	10
F	11
G	1
H	1

At Souther Field the Gosport system of instruction was in use, and the officer in charge of flying kept rating-cards containing a numerical rating of each cadet on each day's performance in the various features of flying.

For the purpose of correlation with the psychological tests, the various numerical ratings of each cadet were

averaged. The range was found to be between 75.00 and 62.80, and the men were placed in fourteen groups as follows (the number of men in each group is indicated):

Group	No. of Cadets
75.00–75.99	1
74.00–74.99	0
73.00–73.99	3
72.00–72.99	9
71.00–71.99	14
70.00–70.99	18
69.00–69.99	11
68.00–68.99	7
67.00–67.99	5
66.00–66.99	2
65.00–65.99	1
64.00–64.99	0
63.00–63.99	0
62.00–62.99	1

2. *Correlations*

The test-scores were correlated with aviation rating by the use of the standard Pearson-Bravais 'product-sum' formula. The coefficients are shown in the following table:

TABLE V

CORRELATIONS BETWEEN TESTS AND AVIATION RATING

	Taylor	Souther
Parabolic Curves: Consistency of Judgment (MV/M)	.139* ± .038	−.013 ± .036
Parabolic Curves: Deviation from Mean Estimate (x/σ)	.109* ± .038	.253* ± .035
Moving Objects: Penalty Score	.232* ± .096	.224* ± .085
Complex Reaction: Time	.170* ± .067	.039 ± .081
Complex Reaction: Errors	.181* ± .066	−.091 ± .078
Complex Reaction: Time Corrected for Error	.256* ± .064	— —
The Maze: Time 3d In and Out	.105 ± .09	.124 ± .079
The Maze: Time 4th In and Out	−.008 ± .09	.020 ± .080
The Maze: Immediate Practice-Effect	−.015 ± .09	.006 ± .081
Dynamometer: Endurance of Pull	−.161 ± .09	.148 ± .098
Dynamometer: Strength of Pull	−.034 ± .09	−.084 ± .09

The signs of the coefficients uniformly agree with + for merit in the test: *i.e.*, a positive correlation with "Errors," in Complex reaction, means that merit, with respect to accuracy, has a positive correlation with aviation rating.

All of the coefficients are low, but some of them (indicated

by an asterisk) are over or near three times their probable error and constitute evidence for correlation. The coefficients falling below this statistical requirement do not indicate the absence of correlation, since in the nature of the case coefficients of genuine but small correlations cannot meet the statistical requirement. The correlation between some of the tests and aviation rating is about as close as some published correlations between Breadth and Length of Skull in Bavarian peasants ($r = .28 \pm .06$), in French peasants ($r = .13 \pm .09$), or in Aino women ($r = .18 \pm .08$) (Davenport: 'Statistical Methods').

The correlation between the total results of the tests and aviation rating was found by the Grade and Rank methods, in which Pearson's standard '*R*' and '*Rho*' formulæ were used:

TABLE

CORRELATION BETWEEN TOTAL TEST RESULTS AND AVIATION RATING

	Taylor Field	Souther Field
Grade Method	$R = .224$; $r = .375 \pm .095$	$R = .124$; $r = .216 \pm .087$
Rank Method	$\rho = .355$; $r = .370 \pm .087$	$\rho = .270$; $r = .282 \pm .077$

The probable errors indicate that the coefficients of correlation (r) are fairly reliable. The relationship between the test-results as a whole, and aviation rating, is about as close as that between stature and length of forearm (.37), or between the statures of father and daughter (.360) or mother and daughter (.284).

With respect to the value of these psychological tests for the detection of aviational ability, our results are as clear as the limitations of the research would permit. The tests, as a whole, and some of them singly, are to some extent diagnostic. Their precise value, however, can be known only after trial under more favorable conditions. The special disabilities of this research may be stated as follows:

1. The test-scores are not good measures of the cadet's ability in the test-performances; they include a large amount of adventitious variability associated with changes in method incident to the early period of learning, and with practice-effect of varying rates; and they are based on a small number of trials wholly limited to this unfavorable period of learning.

2. The number of cases was too small. A correlation table of 50 or 70 is incapable of showing a reliable correlation if the coefficient be small. We should have had at least 200 cases from each field. (Test-results were obtained from many flyers for whom no aviation rating was available.)

3. The aviation ratings are not good measures of aviation ability; they are weighted with military, personal, social, and other qualifications. Also, the factors considered and the method of expressing rating differed sufficiently at the two fields to preclude either the aggregation, or the close comparison, of the results from them.

4. No ratings of the more specific aviational abilities were available for evaluating the respective tests as diagnostic of those special abilities.

5. The range of aviational ability was too narrow to furnish a good criterion of diagnostic tests.

These disabilities were unavoidable under the conditions existing at the time the research was carried out, and no criticism is implied. They should, however, serve as guides in further research designed to find the best psychological tests diagnostic of aviational ability.

Conclusion

In view of the disabilities under which this research has been conducted, the positive results are more important than the quantitative statements indicate. The following tests, in the order written, show promise as a means of detecting aviational ability:

1. Judgment of relative speeds of moving objects,
2. Judgment of parabolic curves,
3. Complex reaction time.

They deserve further trial under conditions suggested by the disabilities listed above, in order that their precise individual value may be determined. Their full diagnostic value could then be learned by finding the correlation between the rating in aviation and an aggregate score in which each test is weighted in proportion to the individual coefficients of correlation between the respective tests and aviation rating.

THE RELATIVE VALUE OF GROUPED AND INTERSPERSED RECITATIONS

BY E. B. SKAGGS, M.S.

Ohio Wesleyan University

The following experiment and discussion is largely an outgrowth of the experiments performed by a number of workers along the line of recitation work in learning. In particular it is an outgrowth of the work done by Witasek[1] and later A. I. Gates.[2] The following work assumes the soundness in general of the conclusions reached by Gates that recitation (attempted recall) is more valuable in learning than mere reading. During the summer of 1919 the writer did a little work, in the foot-steps of Witasek, that showed not only the superiority of the recitation method but indicated that there was an *optimum* combination of readings and recitations. However the work along this line was discontinued.

The purpose of the following work is to throw some light if possible upon this question; granting the superiority of *L*'s plus *R*'s (throughout we will let *L* stand for reading and *R* for recitation) over mere *L*'s, is it better to intersperse *L* and *R* or use some degree of grouping? Put in more practical form we might state the question thus: Is it better for the student to read through his English-German vocabulary, poem or dates or what not, and then try to recall, or should he read two or more times and then try to recall two or more times? In other words shall he intersperse or group? Is there any law indicated? It is upon this question that we have hoped to throw some light—at least we have tried to find indications of some principle or law.

[1] Witasek, S., 'Ueber Lesen und Rezitieren in ihren Beziehungen zum Gedächtnis,' *Zsch. f. Psychol.*, 1906, 44, 161–185, 246–282.

[2] Gates, A. I., 'Recitation as a Factor in Memorizing,' *Arch. of Psychol.*, No. 40, Sept., 1917.

Notwithstanding the good work that has been done by Gates, Witasek and others, the writer wishes to point out a fact that is deserving of some criticism. A recitation (*R*) means not a recitation in a pure sense as used by these investigators, but really means *R* plus *L* (*L* standing for the promptings which, in whatever form given, are after all simply readings). So they have not compared mere *L* with mere *R*, but with a mixed method which complicates the total process a great deal.

Suppose that *S* (subject) takes 6 *L*'s and 6 *R*'s (in the sense used by these workers). In *R* he gives as much as he can and is prompted on the rest. So it goes through the other readings. To take some data from Witasek to illustrate. Keeping the number of original *L*'s constant, that is six in both cases he gives the *S* 5 additional *L*'s and 0 *R*'s. This makes a total of 11. In the next experiment he gives 0 additional *L*'s and then has *S* give 5 *R*'s, again a total of 11. But the two 11's are not strictly comparable—in the first case he has 11 readings. In the second 6 readings plus 5 recitations involving reading in themselves.

It may be replied that while the above procedure complicates the factors involved, nevertheless it lies nearer to actual conditions of learning. The student looks away from his book, reciting all he can, and when stopped, takes a look (a prompt) and so goes on. From this standpoint the use of the above mentioned mixture is probably justifiable. Certainly it complicates matters already sufficiently complicated if we desire to analyze out essential processes. We wish merely to indicate that these workers were not comparing a method of learning without recitation with a method of learning with recitation in any strict sense.

We have tried to eliminate this mixture in Part I. and II. of our experiment, by keeping an *L* an *L* and an *R* an *R* in a strict sense. No promptings of any kind have been given. When it is stated that there were 6 *R*'s grouped, *e.g.*, we mean *S* gave all he could, then again tried to give all he could. And so on for six times. In discussing the imprinting and recall factors we believe that our method permits of more

clean-cut comparisons than would have been the case if we had used a more mixed method. However, in Part III. of our experiment we have used '*R*' in the mixed sense.

Material.—Part I. of the investigation uses nonsense syllables for the learning material. A series of twelve such syllables constituted a learning series. Part II. and III. use sense material in the form of poetry, eight-line stanzas from Byron's 'Don Juan' furnishing the material.

Part I. Nonsense Material

(*L* + *R Used in Strict Sense of Term*)

Apparatus.—A modified Wirth memory apparatus was employed. A large screen of black cardboard completely hid the experimenter from the subject. A slit was made near the center of the screen so that the window of the machine coincided with it. Thus *E* could operate the machine, take notes, etc., without *S* knowing what was happening. Slips of white typewriting paper with the syllables typewritten thereon in capital letters were used. A metronome, in a felt rubber padded box across the room and hidden from the subject, beat the time. The ticking of this instrument was barely audible and was very seldom heard by *S*.

Method.—Twelve syllables (constituting a series) were shown in succession through the window in the Wirth machine by means of the rotating drum. The tempo selected was one syllable per $1\frac{1}{4}$ second. This time was selected after some exploring for a suitable tempo. We wanted a time that would not "rush" *S* or be uncomfortable, and on the other hand avoid exposure time long enough to put a premium on anticipation, repetition, and attention relaxations. Finally the above time was selected as suiting our purposes fairly well. As there were always two blank spaces on the rotating drum, in every learning there was a period of two and one half seconds during which nothing was shown and before the second or next reading could begin.

The following instructions were emphasized to each subject:

1st. In learning (reading) never anticipate, *i.e.*, reach out for, try to get before it comes into view, the next syllable. Simply take the syllable given, hold it as attentively as possible, and then take the next one as it appears.

2d. In reading, say the syllable but once. Never repeat to make it stick.

3d. Try to keep the syllables as meaningless as possible. Do not search for some meaning, do not put meaning into the syllable.

4th. Report all distractions and any conditions that would make the results non-comparable with other records.

5th. In recitation, do your best. Do not give up if few syllables come. Keep on trying.

Four methods were used in the course of the experiment:

I. The Interspersed Method (6*L* and 6*R*).

II. *b*. The "Slightly Grouped" Method (2*L*–2*R*–2*L*–2*R*–2*L*–2*R*).

II. *a*. The "Intermediate Method" (3*L*–3*R*–3*L*–3*R*).

II. The Extreme Grouped Method (6*L*–6*R*).

In every method used there were, as a total, 6*L*–6*R*, making in every case twelve. At the end of each experiment there was given an immediate reproduction test. One hour later the subject again recalled all possible. Twenty-four hours later *S* again tested himself or herself to see what could be recalled.

As there are twelve syllables in each series, we worked out averages on this basis. Thus if *S* reproduces, say, 11.5 syllables as his average immediate reproduction, this is with reference to the perfect score, namely 12. In this case, if one hour later the average amount recalled was 7 syllables, *e.g.*, then the per cent. of retention would be 65.86. There is no part credit given—the material is right or wrong! Each syllable is one unit of the perfect score 12.

The data following were obtained from nineteen university students in the elementary psychology course and from five trained subjects—subjects[1] who had been given previous

[1] We wish to express thanks to the trained graduate persons who, as subjects, did their work conscientiously and whose introspections were essential—Miss E. Gordon, Mrs. E. B. Skaggs, Miss M. Gurnsey, and Miss S. Mebai.

training in learning nonsense syllables and four of whom were graduate students in psychology. (The writer is included in this list.) It is chiefly upon the work and introspections of these five trained subjects that we wish to base any finer distinctions, although each is used as a check on the other group.

TABLE I

SHOWING INDIVIDUAL AVERAGES AND GENERAL AVERAGES FOR GROUP I. USING THE INTERSPERSED METHOD (METHOD I, *6L–6R* INTERSPERSED)

Subject	Amt. Immed. Reprod.	Amt. Hour Recall.	Amt. 24 Hr. Recall	Per Cent. Retain. 1 Hour	Per Cent. Retain. 24 Hours	Errors Immed.	Hour	Day
1. E.B.C.....	11.5	7.0	6.75	65.86	64.51	.75	1.5	1.5
2. M.D.....	11.25	10.75	10.75	93.75	93.33	.25	.75	.5
3. D.D.....	9.5	7.75	7.30	81.59	74.05	1.75	2.25	1.00
4. R.R.....	11.25	11.25	10.00	97.72	87.12	.75	.50	1.50
5. C.W.....	10.5	10.50	10.50	100.00	100.00	.50	.75	.75
6. G.F.P.....	8.75	5.75	4.50	63.65	54.98	3.00	1.50	1.00
7. C.W.C.....	8.00	8.00	6.50	89.10	83.54	1.60	2.25	2.25
8. R.T.D.....	12.00	12.00	11.00	100.00	96.16	0.00	0.00	0.00
Totals.....	82.75	73.00	67.30	691.67	651.69	8.60	9.50	8.50
Average.....	10.34	9.13	8.41	86.50	81.46	1.08	1.18	1.06
M.V.[1] =.....	1.21	2.00	2.15					

TABLE II

SHOWING INDIVIDUAL AVERAGES AND GENERAL AVERAGES FOR GROUP I. USING THE EXTREME GROUPED METHOD (METHOD II, *6L–6R* GROUPED)

Subject	Amt. Immed. Reprod.	Amt. Hour Recall	Amt. 24 Hr. Recall	Per Cent. Retain. 1 Hour	Per Cent. Retain. 24 Hours	Errors Immed.	Hour	Day
1. E.B.C.....	7.00	6.00	5.50	83.93	77.86	1.00	.75	.75
2. M.D.....	7.50	7.50	7.50	97.50	96.25	.25	.75	.75
3. D.D.....	8.00	6.50	5.50	85.62	77.13	.75	1.50	.25
4. R.R.....	9.00	8.50	8.50	94.44	94.10	.75	1.75	1.75
5. C.W.....	8.25	8.00	7.25	96.43	87.74	.25	.25	.00
6. G.F.P.....	7.75	5.75	4.50	72.29	55.83	2.00	2.00	1.25
7. C.W.C.....	7.80	7.33	4.50	89.77	66.66	.60	2.00	3.25
8. R.F.D.....	7.00	6.25	5.66	84.92	73.81	1.75	1.75	1.00
Totals.....	62.30	55.83	48.91	704.90	629.38	7.35	10.75	9.0
Average.....	7.79	6.95	6.10	88.10	78.67	.93	1.33	1.13
M.V. =.....	0.48	0.85	1.22					

[1] In all cases our M.V. is taken from the mean.

TABLE III

SHOWING INDIVIDUAL AND GENERAL AVERAGES FOR GROUP II. USING THE INTERSPERSED METHOD

Subject	Amt. Immed. Reprod.	Amt. Hour Recall.	Amt. 24 Hr. Recall.	Per Cent. Retain. 1 Hour.	Per Cent. Retain. 24 Hours	Errors Immed.	Hour	Day
9. H.S.	11.37	11.66	11.37	100.00	100.00	.30	.30	.70
10. R.H.M.	9.70	8.70	8.70	89.70	89.70	.00	.70	.70
11. A.S.	11.50	10.50	8.50	90.83	73.74	.50	.25	.25
12. E.G.	11.30	11.00	10.30	97.00	94.67	.00	.30	.70
13. D.S.	11.00	9.70	7.50	87.52	70.90	.70	.30	.30
17. E.S.	8.00	6.70	7.00	82.60	86.80	1.00	1.30	1.30
18. M.S.	9.70	9.00	9.70	93.88	100.00	1.00	1.00	1.00
19. G.B.	10.70	9.00	8.70	85.90	75.90	1.00	1.00	1.00
20. J.L.	11.00	10.30	9.30	91.30	82.30	.70	1.30	1.30
21. F.F.	11.00	8.00	8.00	72.27	71.91	.30	1.70	1.00
22. C.M.	10.00	9.7	9.30	96.30	92.70	1.30	1.70	1.70
Total	115.27	104.26	98.37	987.30	938.62	6.80	9.90	10.0
Average	10.57	9.5	8.95	89.75	85.33	.6	.9	.9
M.V. =	0.79	1.09	.96					

TABLE IV

SHOWING INDIVIDUAL AND GENERAL AVERAGES FOR GROUP II. USING METHOD II. *a* (INTERMEDIATE—3*L*-3*R*-3*L*-3*R*)

Subject	Amt. Immed. Reprod.	Amt. Hour Recall.	Amt. 24 Hr. Recall	Per Cent. Retain. 1 Hour	Per Cent. Retain 24 Hours	Errors Immed.	Hour	Day
9. H.S.	12.00	12.00	12.00	100.00	100.00	.00	.00	.00
10. R.H.M.	11.00	10.00	10.00	87.00	87.00	.00	1.00	1.00
12. E.G.	11.00	10.30	11.00	94.00	100.00	1.00	1.30	1.00
13. D.S.	10.30	9.00	7.70	86.52	73.73	.30	1.30	1.30
17. E.S.	8.00	8.50	7.30	100.00	92.00	1.00	.70	1.00
18. M.S.	9.70	8.00	7.30	82.87	72.22	.30	.00	.30
19. G.B.	11.30	9.70	9.70	84.80	85.00	.30	1.70	1.30
20. J.L.	11.70	11.70	10.70	100.00	91.00	.30	.30	.30
21. F.F.	8.00	9.00	7.50	100.00	87.85	.00	.00	.00
22. C.M.	8.00	7.50	7.70	90.00	96.66	1.30	1.30	1.70
Totals	101.00	95.70	90.90	925.20	885.46	4.50	7.60	7.9
Average	10.10	9.57	9.09	92.52	88.55	.45	.76	.79
M.V. =	1.34	1.17	1.59					

TABLE V

SHOWING INDIVIDUAL AND GENERAL AVERAGES FOR GROUP II. USING THE EXTREME GROUPED METHOD (*6L–6R*)

Subject	Amt. Immed. Reprod.	Amt. Hour Recall.	Amt. 24 Hr. Recall.	Per Cent. Retain. 1 Hour	Per Cent. Retain. 24 Hours	Errors Immed.	Hour	Day
9. H.S.	9.33	9.66	9.70	100.00	100.00	.30	.30	.30
10. R.H.M.	7.30	8.00	7.50	100.00	93.00	1.30	1.00	1.00
11. A.S.	11.00	10.00	6.25	90.86	74.69	1.00	.75	1.00
12. E.G.	8.60	9.00	—	100.00	—	.70	1.00	—
13. D.S.	10.00	8.50	7.30	88.88	74.50	1.00	1.50	.70
17. E.S.	8.00	7.70	7.30	91.70	88.40	.70	.70	.70
18. M.S.	10.00	8.00	7.00	76.97	67.27	.30	1.00	1.00
19. G.B.	9.70	9.00	8.30	88.30	85.60	.70	.70	1.30
20. J.L.	10.70	11.30	10.70	100.00	94.00	.30	.30	.70
21. F.F.	7.70	7.00	7.00	92.60	92.60	.70	1.00	1.00
Totals	93.00	88.00	71.00	929.31	770.06	7.00	7.75	7.70
Averages	9.30	8.80	7.78	92.93	85.56	.70	.77	.85
M.V. =	1.05	.98	1.07					

TABLE VI

COMPARING AVERAGES FOR GROUP II., IN METHODS I., II., AND II. *a*

Method.	Amt. Immed. Reprod.	Amt. Hour Recall.	Amt. 24 Hr. Recall.	Per Cent. Retention 1 Hour	Per Cent. Retention 24 Hours
I. *6L–6R* Intersp.	10.57	9.50	8.95	89.75	85.33
II. *a* $3L_3R_3L_3R$ Intermed.	10.10	9.57	9.09	92.50	88.55
II. *6L–6R* Grouped	9.30	8.80	7.78	92.93	85.56

TABLE VII

COMPARING RESULTS (UNTRAINED SUBJECTS, GROUP I.) METHOD I. AND II., ACCORDING TO INDIVIDUAL DISTRIBUTION

Retention Period	Number Better in Method I.	Number Better in Method II.	Number Indifferent	Total
Immediate	8	0	0	8
Hour	7	0	1	8
24 hours	7	0	1	8

TABLE VIII

COMPARING RESULTS (UNTRAINED SUBJECTS, GROUP II.) METHOD I. AND II., ACCORDING TO INDIVIDUAL DISTRIBUTION

Retention Period	Number Better in Method I.	Number Better in Method II.	Number Indifferent	Total
Immediate	8	1	1	10
Hour	7	2	1	10
24 hours	7	2	0	9

TABLE IX

COMPARING RESULTS (UNTRAINED SUBJECTS, GROUP II.) METHOD I. AND II. *a*, ACCORDING TO INDIVIDUAL DISTRIBUTION

Retention Period	Number Better in Method I.	Number Better in Method II. *a*	Number Indifferent	Total
Immediate	4	4	2	10
Hour	4	6	0	10
24 hours	3	7	0	10

TABLE X

COMPARING RESULTS (UNTRAINED SUBJECTS, GROUP II.) METHOD II. *a* AND II., ACCORDING TO INDIVIDUAL DISTRIBUTION

Retention Period	Number Better in Method II. *a*	Number Better in Method II.	Number Indifferent	Total
Immediate	7	1	1	9
Hour	8	0	1	9
24 hours	6	0	2	8

TABLE XI

GIVING AVERAGES FOR FOUR METHODS, I., II. *b*, II. *a*, AND II. FOR TRAINED SUBJECT, E.B.S.

Method	Amt. Immed. Reprod.	Amt. Hour Recall	Amt. 24 Hour Recall	Per Cent. Retain. 1 Hour	Per Cent. Retain. 24 Hours	Errors Immed.	Hr.	Day
I.	11.70	11.30	9.50	95.92	81.56	.20	.30	2.20
II. *b*	10.30	9.30	7.80	90.00	75.60	.00	.70	1.20
II. *a*	9.80	9.30	7.70	95.15	78.27	.50	.50	2.50
II.	6.20	6.20	5.20	100.00	84.75	.70	.70	.80

TABLE XII

GIVING AVERAGES FOR TWO METHODS, I., INTERSPERSED AND II., GROUPED FOR TRAINED SUBJECT, S.M.

Method	Amt. Immed. Reprod.	Amt. Hour Recall	Amt. 24 Hour Recall	Per Cent. Retain. 1 Hour	Per Cent. Retain. 24 Hours	Errors Immed.	Hr.	Day
I. (6*L*–6*R*) Intersp.	11.70	11.30	9.00	97.10	79.00	.30	.30	1.00
II. (6*L*–6*R*) Grouped	9.50	9.50	7.10	98.60	77.60	.50	.70	1.00

TABLE XIII

GIVING AVERAGES FOR FOUR METHODS, I., II. *b*, II. *a*, AND II. FOR TRAINED SUBJECT, M.G.

Method	Amt. Immed. Reprod.	Amt. Hour Recall	Amt. 24 Hour Recall	Per Cent. Retain. 1 Hour	Per Cent. Retain. 24 Hours	Errors Immed.	Hr.	Day
I.	11.3	10.3	7.7	91.13	66.89	1.0	1.3	2.3
II. *b*	11.3	10.0	10.0	87.20	87.20	.3	1.3	1.3
II. *a*	10.3	9.3	8.3	84.54	78.18	1.7	.3	.7
II.	9.0	10.0(?)	7.5	100.00?	75.24	.3	.3	1.7

TABLE XIV

GIVING AVERAGES FOR FOUR METHODS, I., II. *b*, II. *a*, AND II. FOR TRAINED SUBJECT, E.I.G.

Method	Amt. Immed. Reprod.	Amt. Hour Recall	Amt. 24 Hour Recall	Per Cent. Retain. 1 Hour	Per Cent. Retain. 24 Hours	Errors Immed.	Hr.	Day
I.	12	11.7	11.3	97.22	85.83	0	0	0
II. *b*	11	11.3	11.0	100.00	95.80	0	0	0
II. *a*	10	9.3	7.5	92.60	72.20	.7	1.3	2.5
II.	8	7.7	6.3	92.19	77.08	1.3	1.7	1.5

TABLE XV

GIVING AVERAGES FOR FOUR METHODS I., II. *b*, II. *a*, AND II. FOR TRAINED SUBJECT, I.D.S.

Method	Amt. Immed. Reproduced	Amt. Hour Recall	Amt. 24 Hour Recall	Per Cent. Retain 1 Hour	Per Cent. Retain. 24 Hours
I.	11.50	11.50	10.30	100.00	75.00
II. *b*	10.75	10.25	9.70	93.15	87.27
II. *a*	9.00	8.50	8.00	93.93	90.00
II.	6.00	5.50	4.75	90.84	78.33

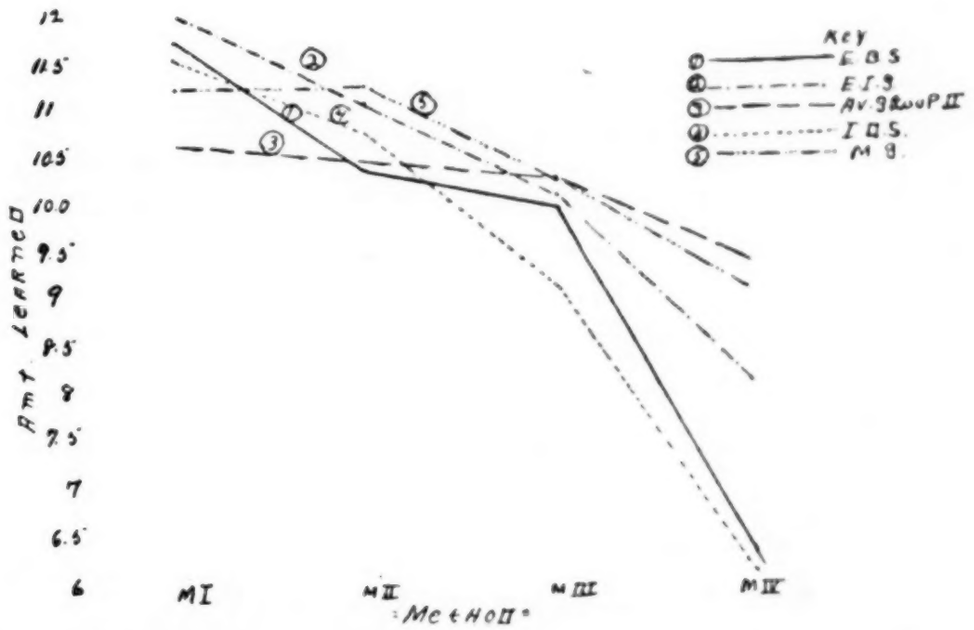
RESULTS FROM UNTRAINED SUBJECTS

Considering first the interspersed and extreme grouped methods, we find our results clean cut. The grouped method is very inferior in comparison. Tables I. and II. (giving the results from group I., untrained subjects) show a marked difference in favor of the interspersed method in the amount learned, the hour and the twenty-four hour retention. Group II. likewise show the same difference though it is slightly less marked (see Tables III., V., and VI., also Graph I.).

In Tables VII. and VIII. we have the same facts shown in a somewhat different way. Here we have taken the indi-

viduals who did better by one or the other method and so recorded the fact, regardless of the amount of superiority or inferiority. In group I. no one does better by the grouped method in the amount learned (Table VII.). Of the ten people in group II., only one did better by the grouped method, one being indifferent.

When we turn to the percentage of retention in the two methods, we do not find any significant differences. Though one method may give but few learned, yet is it not possible



GRAPH I. Curves showing relative amounts learned by various methods. (Four trained subjects and average of Group II., untrained.) Nonsense syllables.

that what little is learned will stick longer? To test this we took as standard the amount learned and taking the amount retained one hour and twenty-four hours later, worked out the percentage. Our question seems to be answered in the negative, there being no regularity in the results. As to errors in the two methods, we again find our results very irregular.

Group II. used, besides the two above mentioned methods, an intermediate method, 3*L*, 3*R*, 3*L*, 3*R*. Tables VI., IX., and X. give the comparative results. As we might expect, this method stands intermediate in amount learned. The

curve in Graph I. shows this very well. However, in the 24-hour retention the method actually is better than the interspersed. Table IX. indicates that it seems not to matter whether the interspersed or mildly grouped method is used. Four do better in the interspersed method, four better in the intermediate method, and two are indifferent.

Table X., comparing the mildly grouped with the more extreme grouped method shows a marked superiority of the mildly grouped method.

The above data indicate that as one goes from the interspersed method to more and more grouped method, the efficiency of the learning decreases and that a point is quickly reached where the grouped method shows a marked inferiority.

Results from the Five Trained Subjects

It is the results of these five trained subjects that we wish most to emphasize; (1st) because each was given previous preliminary training, (2d) because of their conscientious work and careful following of the conditions of the experiment, and (3d) because of the value of their introspections.

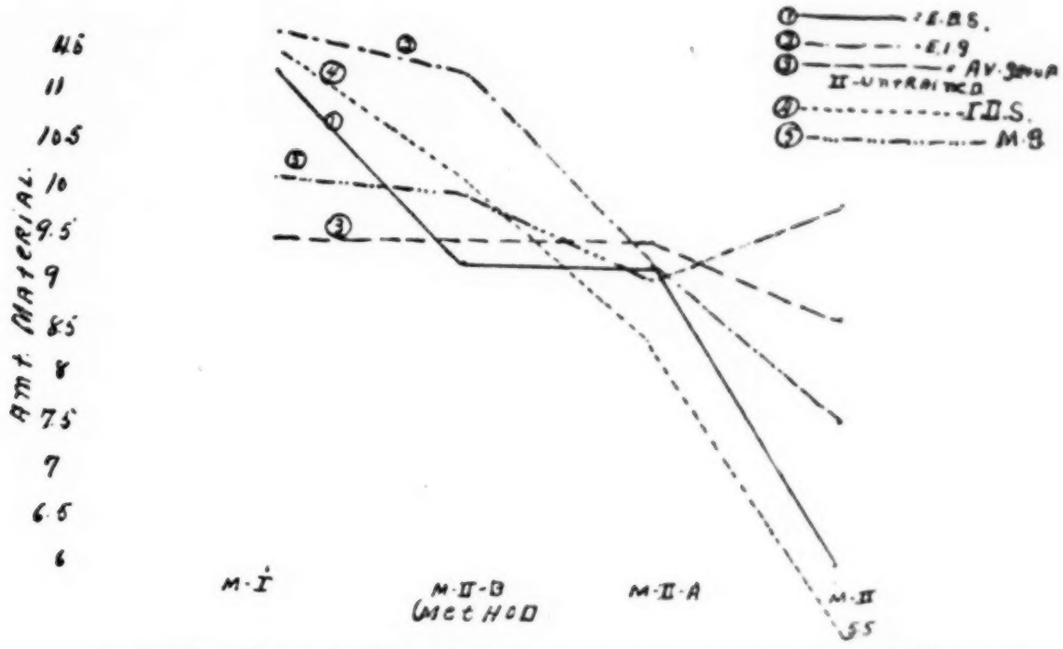
The records of the trained subjects indicate the principle stated above better than do the untrained subjects. Graphs I., II., and III., give curves indicating the principle. (S.M. has no curve plotted because she participated in only two methods.) Tables XI. to XV. give the data tabulated.

A preliminary training was given to all five subjects before any records were kept—the training ranging from ten to fifteen series learned. The averages are based on from four to six learnings in each method, after weeding out learnings which involved abnormal conditions, as fatigue, distractions, etc.

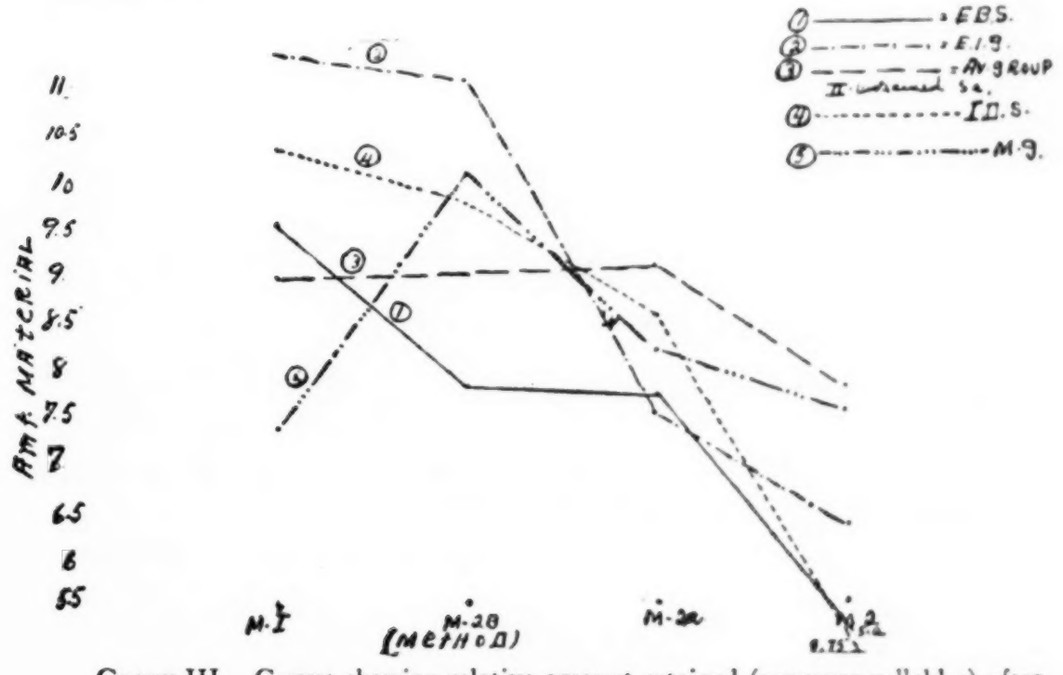
According to the results obtained by the trained subjects the following principle seems strongly indicated: interspersed learning and recitation is better than any form of grouping, the efficiency decreasing as the grouping increases.

Consideration of factors involved which seem to be more effective in the interspersed than in the grouped methods.

Under the conditions of this experiment it is clear that,



GRAPH II. Curve showing relative amount retained (nonsense syllables) after an interval of one hour.



GRAPH III. Curves showing relative amount retained (nonsense syllables) after twenty-four hour interval.

either the interspersed method must involve helping factors not found at all in the grouped methods or else utilize the factors to a greater degree. On the basis of introspective reports we hazard to suggest the following factors as operating in favor of the interspersed method.

1. Time and again our subjects say that in the interspersed method they can find their weaknesses and check up thus on the next learning fixing these eluding syllables. This is doubtless a superficial expression of the more fundamental facts. However, the essential fact seems to be this: Certain syllables are but partly secured, the subject is not certain of them. This furnishes an incentive or attitude whereby, on the next learning, there is a higher degree of concentration and a determination to get the syllables that makes for efficiency. Again, the fragmentary part of doubtful syllables constitutes a mental set favorable to the true syllable when it comes—gives it an extraordinary attention value.

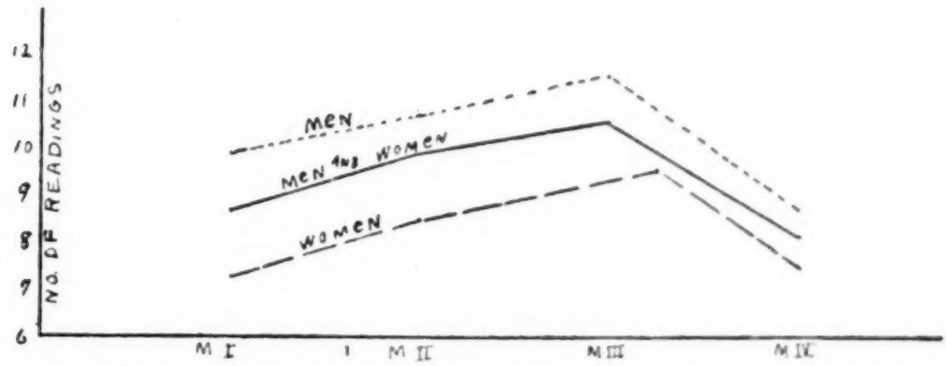
2. Each time the subject recites he repeats (without the stimulus) the syllable. As a mere matter of habit, this tends to fix the syllable, and so *S* does not pay much attention to this or that one next reading but concentrates on the 'new' or 'wrongly recited ones.' This is another condition expressed by practically every subject. It is clear that this method can be little utilized in the grouped methods. You cannot check up on your weaknesses so well here, you cannot have that satisfaction of knowing you have the syllable thoroughly and so gauging your line of attentive attack on new materials. But this repeating of the syllable without the stimulus being before you and at leisure seems to be a very important factor.

Gates finds that his subjects had ill success in forming associations by a mere shifting of attention from one to another contiguous element. This leads Gates to hold as insignificant the importance of attention shift from element to element contiguously. However, there is something else we must consider. We would say that the very core of association forming is the shift of attention from one element to another. But this attention must be of active nature, involving 'anticipation,' a 'reaching out for' the next element, to be effective. Under this condition, when the syllable comes a bond seems to bind the contiguous elements. True, the mere contiguity in consciousness of the two syllables is little effective.

Just what constitutes this anticipation, this reaching out, we are not sure. But all can experience it—all do experience it in memory recall and all can experience it in slow exposures of nonsense syllables. Syllable A comes. If I have seen it before and have time, I definitely employ this anticipation. It seems to be a sort of active attitude, a more or less complete recall into consciousness of the next syllable.

Certainly it involves partially at least the act of recall. The syllable flashes into consciousness before the actual stimulus comes. On the other hand, the anticipation may not thus recall the syllable and one gets something akin to the lowest form of recognition. The association bridge is almost crossed but the experience does not quite come.

Now in recitation, this active attitude, this anticipation, can operate as it cannot in reading nonsense syllables. Reading discourages this process and the more we group our readings the less this factor operates. In so far as this anticipation process enters at all into learning (reading) it is no longer a reading process! The very core of recitation is this anticipation.

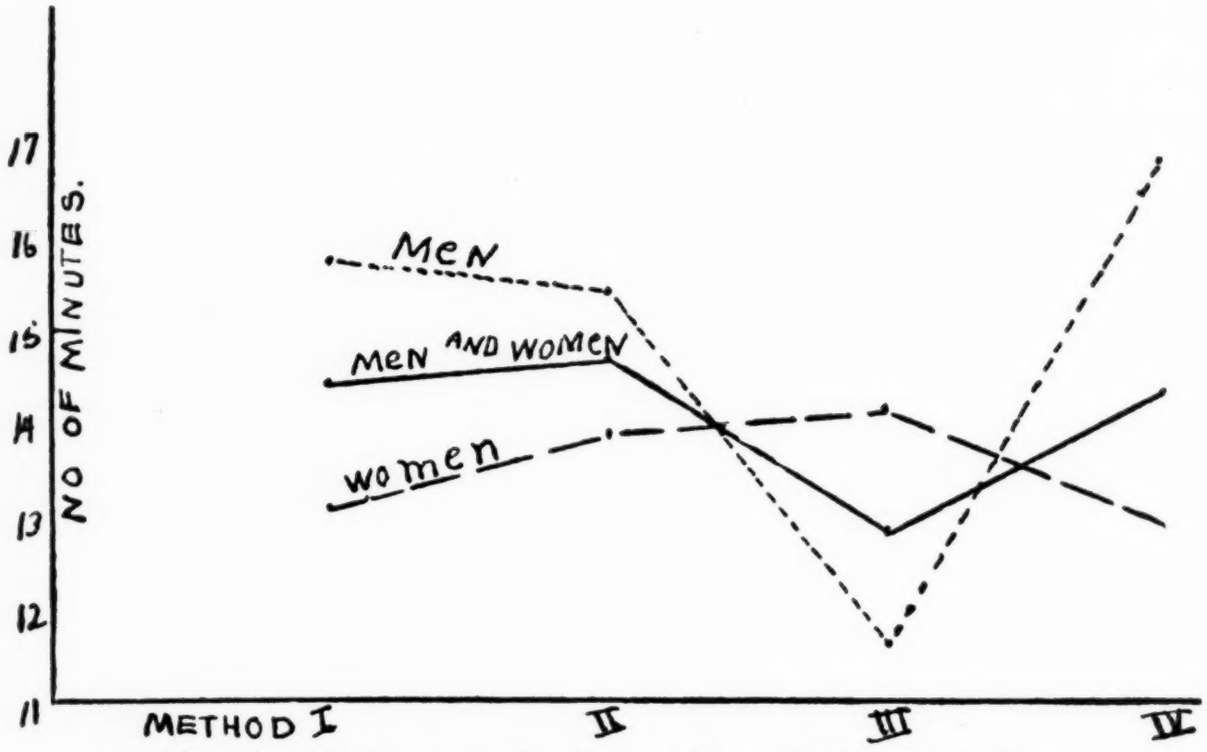


GRAPH IV. Showing comparative efficiency of four methods as to number of L's required to learn.

3. The 'setting time' seems better utilized in recitation and the consequent different types of inhibitory processes are less destructive. Often the subject reported that he had a syllable until the next was shown and then the former was blotted out of existence. It was a help to slide over the syllables and hold longer, in consequence, to the troublesome syllable. The interfering tendencies of retroactive inhibition

were thus decreased. Something of the same factors found in divided rather than concentrated repetitions in memorizing a poem, seems to be effective here.

It is clear that grouping the readings together does not give these above mentioned factors as much opportunity to be effective as the interspersed. The same general principle that makes reading plus recitation better than mere reading alone,



GRAPH V. Showing comparative efficiency of four methods as regards time.

and that makes divided repetitions better than concentrated, makes interspersed reading-recitation better than any form of grouped method.

INCIDENTAL RESULTS AND SUGGESTIONS

1. Using the products moment formula (as there were a number of tie cases we did not think it best to use the rank method), the coefficient of correlation was worked out for the 24 subjects between amount learned and the amount retained after 24 hours. We found r to be .65, probable error as .078.

This is in keeping with the general trend of findings, namely, that those who learn best retain longest.

2. Using our methods of presentation, learning and scoring, our curve of forgetting drops but very little in 24 hours. On the average about 10 per cent. of the syllables were lost during the first hour and then about 5 to 10 per cent. is lost at the end of 24 hours. The main suggestion from our curve is this: we should speak of learning curves rather than any one learning curve. To hold Ebbinghaus' curve as typical seems hardly fair to the facts. The kind of forgetting curve one gets depends on such things as kind of material used, method of learning, method of testing, etc.

3. The fact that some subjects did as well on the extreme grouped readings as on the more interspersed, and the fact that other subjects were very helpless indeed in the extreme grouped methods, seemed to indicate two if not types, then extremes of a distribution curve, as regards learning. One group might be called a 'mechanical sort' the other an 'associative group.' (We do not wish to be charged with calling these 'types.') Subjects A.S., D.S., M.S., J.L., *e.g.*, (Table V.) illustrate that group in which reading alone gives almost as good results as reading plus recitation. On the other hand subjects E.B.S., I.D.S., M.G. seem helpless in the method of grouped readings.

These variations need further investigation and might throw some interesting light on differences in learning.

Part II. Sense Material (Poetry)

(*L and R Used in Strict Sense of Term*)

As the first part of the investigation showed rather clean cut results it was thought well to extend the experiment to include sense material. After all, from the practical standpoint, we wish to know which learning method is best as applied to sense material. Other investigators (Gates, *e.g.*) found that with sense material his principles held less true.

Consequently eight line stanzas of poetry were selected (Byron's 'Don Juan'). As in Part I., the *L*'s and *R*'s were used in the strict sense, no promptings being used whatever in the recitations. An *L* means that the subject read through

the stanza from beginning to end with concentrated attention, while an *R* means an attempted recall without any prompting.

Thirty-six persons acted as subjects, all but two undergraduate students taking elementary psychology. Nineteen of the subjects were men and 17 were women.

Procedure.—The stanzas were typewritten upon slips of plain white paper. In *L* the whole method was used. The subjects sat about tables in the room with very little disturbance going on at any time. Each subject took the number of stanzas he thought he would use on that occasion and started work. For the most part an hour's work was the maximum, though in a few cases a longer work period was granted. The subjects were instructed to stop learning at once if they began to feel at all bored, tired, or fatigued. A general supervision was kept over the subjects as they thus worked.

Four methods were used here, each method involving two preliminary but unrecorded readings. The methods were:

I. *L R L R L R*, etc., until learned.

II. *LL RR LL RR*, etc., until learned.

III. *LLL RRR LLL RRR*, etc., until learned.

IV. *LLL RRR LL RR L R L R*, etc., until learned.

As can be seen Method I. is the truly interspersed method, II. more grouped, and III. the most grouped. Method IV. is a mixed method involving III., II., and I. in order and continuing with Method I. as indicated above, until learned.

The subject noted the time by his watch, and started to learn immediately. The two preliminary readings were then done and then he continued on his readings and recitations as outlined in the methods above. When he reached the place where he thought he might reasonably recall the poetry correctly he wrote it out (the previous recitations being mental or whispered). He then checked up his written work by a check reading. If there were no mistakes, he recorded the time and this ended this learning. However, there were usually a few minor mistakes, and in this case, subject counted the check reading in his record and then tried to recite the poem again (but only mentally or whispered as previously),

again taking a check reading. This was kept up until a perfect reproduction was made. Then the time was noted. So a record of *L*'s and *R*'s and the time required was kept for each learning.

For the most part four records were secured from each subject for each of the four above mentioned methods (a few, however, did only three). To eliminate practice effects on any method, they were rotated, though at no time was the same method used to begin with (to eliminate the warming up effects hurting any method).

Unfortunately, however, a number did not understand the directions for including method IV. in the rotating plan, waiting until all other methods were finished before doing this method. This, of course, gives Method IV. the advantage of the previous practice. These cases have been indicated in the tables by placing an asterisk at the end of the row. Whenever these have been included in the comparisons, it has been noted.

The average number of *L*'s, *L*'s and *R*'s, and time is given in the following tables for each subject. On the basis of these the general averages are worked out. Due to the fact that the number of records for each method for any one subject is so limited, the general averages probably give truer results than any method of comparing by using absolute individual standings.

An inspection of the tables and graphs shows that, with regard to number of readings required to learn the stanza, Method I. (the interspersed) is best of the three. Two is next best and three takes most readings. If we consider Method IV. (the mixed method) also, then it stands first. The above is true for the averages of the men and women taken separately as well as for their combined averages. (See Graph IV.)

As regards the time required (see Graph V.), Method III., the most grouped method, is slightly the best. However, the differences in time are very small at most and probably insignificant.

Summary.—Using sense material and comparing the three methods used (the interspersed, the slightly grouped, and the more grouped methods) we find that,

TABLE XVI

SHOWING COMPARATIVE RESULTS FOR METHODS I., II., III., AND IV. FOR 19 MEN

Subject No.	Method I.			Method II.			Method III.			Method IV.		
	L's	*LR*'s	Time	*L*'s	*LR*'s	Time	*L*'s	*LR*'s	Time	*L*'s	*LR*'s	Time
1.....	12.75	25.50	16.25	13.00	25.50	11.00	12.00	22.75	13.19			
2.....	18.00	36.25	17.25	26.50	53.00	22.00	25.75	51.50	19.50	24.00	48.00	19.00
3.....	7.50	15.50	14.50	7.75	15.25	16.00	7.50	15.75	16.75	3.70	8.30	18.70
4.....	8.50	17.00	22.80	11.70	22.40	28.30	12.25	22.75	26.75	10.30	20.60	20.30
5.....	10.50	21.00	17.77	13.25	25.25	15.00	13.00	24.00	13.00	9.00	17.50	11.50
6.....	9.00	18.00	——	5.50	11.00	——	6.00	12.00	——	5.00	9.50	——
7.....	6.25	12.25	10.50	7.00	12.75	9.88	5.75	11.00	10.50	——	——	——
8.....	4.00	8.00	7.25	5.50	10.00	10.75	8.50	15.50	12.00	——	——	——
9.....	5.30	10.60	10.00	6.00	11.70	11.00	7.30	14.00	12.00	8.00	16.00	17.00
10.....	5.30	11.30	19.70	4.70	10.40	14.70	5.75	12.75	17.25	6.30	12.60	18.70
11.....	13.50	27.00	14.60	13.50	26.25	14.00	15.00	28.00	13.00	12.50	25.00	25.00
12.....	13.75	27.50	15.25	11.50	22.00	14.00	12.00	22.00	11.50	11.50	22.50	11.25
13.....	6.75	13.50	7.25	10.25	19.75	7.33	15.00	27.75	12.25	10.70	20.30	8.00
14.....	11.75	23.50	12.62	13.00	25.50	11.38	12.25	23.50	12.88	9.30	18.70	10.70
15.....	15.75	31.50	18.50	13.75	27.50	17.50	13.50	27.00	16.25	16.30	29.30	14.00
16.....	12.00	24.00	27.00	13.00	25.70	22.00	15.70	31.40	22.00	13.70	27.40	21.30
17.....	8.00	16.00	16.30	9.00	16.70	17.70	10.00	18.30	19.00	10.30	20.30	21.00
18.....	6.70	13.40	14.70	6.70	13.40	14.70	7.00	13.70	15.30	6.70	13.70	16.00
19.....	11.70	23.40	20.00	12.00	23.70	19.30	15.00	29.30	17.70	11.70	23.40	19.30
Total....	187.00	375.20	282.24	203.60	397.75	276.54	219.25	422.95	208.82	149.00	333.10	251.75
Av.....	9.84	19.76	15.78	10.71	20.93	15.40	11.53	22.26	11.60	8.76	18.51	16.78
M.V.....	3.27	6.52	3.82	3.58	7.10	3.93	3.58	6.88	4.17	3.50	7.62	3.90

TABLE XVII

SHOWING COMPARATIVE RESULTS IN METHODS I., II., III., AND IV., FOR 17 WOMEN

Subject No.	Method I.			Method II.			Method III.			Method IV.		
	L's	*L*'s and *K*'s	Time	*L*'s	*L*'s and *K*'s	Time	*L*'s	*L*'s and *K*'s	Time	*L*'s	*L*'s and *K*'s	Time
21	7.00	14.00	11.00	8.25	15.50	14.25	8.75	16.00	11.00	7.00	14.50	10.25*
22	10.50	21.00	13.25	14.25	28.25	14.25	17.25	34.00	15.25	8.00	16.00	11.70*
23	4.75	9.50	10.50	6.25	11.75	9.75	5.50	10.00	10.00	5.50	11.00	6.50*
24	6.75	13.50	12.25	6.25	12.50	10.00	6.25	12.50	11.00	6.30	12.30	12.70
25	7.00	14.00	10.30	10.00	19.70	14.00	10.70	20.40	16.00	9.30	18.60	12.30
26	5.10	10.00	8.50	5.30	10.00	8.00	4.50	9.00	7.95	4.50	8.75	8.25
27	9.30	18.60	14.00	10.00	20.00	13.70	13.00	25.30	11.30	8.00	16.00	14.50
28	8.30	16.60	16.30	9.00	17.75	16.70	14.30	28.60	18.30	10.00	20.00	8.30
29	6.75	13.50	10.75	7.25	14.00	10.25	7.00	14.00	9.90	6.67	13.67	9.70
30	10.30	20.60	15.30	10.30	20.60	16.20	11.30	22.00	17.70	7.30	14.60	11.50
31	6.00	12.00	15.00	7.30	14.30	19.70	9.00	18.00	22.00	7.00	14.00	17.70
32	9.00	18.00	15.70	10.70	21.40	15.70	10.00	19.30	11.30	9.70	19.40	11.30
33	8.00	16.00	17.50	11.70	23.00	25.70	11.30	22.00	26.00	9.00	18.00	20.80
34	4.75	9.50	10.25	7.00	13.25	10.75	7.00	12.70	13.30	7.00	13.30	11.00
35	5.25	10.25	9.25	7.00	13.00	8.50	10.50	19.00	11.00	7.00	13.30	9.30
36	5.30	10.60	9.70	5.87	11.40	8.90	6.50	12.75	9.20	7.00	14.00	11.03
37	9.30	18.60	22.70	7.30	14.60	19.70	8.30	16.60	19.30	8.00	16.00	22.00
Total	123.35	246.25	222.25	143.72	281.00	236.05	161.15	312.15	240.30	127.27	253.40	218.80
Av.	7.25	14.48	13.09	8.45	16.52	13.88	9.48	18.36	14.13	7.49	14.90	12.87
M.V.	1.63	3.23	2.96	1.97	4.12	3.67	2.66	5.15	4.18	1.13	2.31	3.46

(1) as regards the number of readings and recitations, the interspersed method is the best, the slightly grouped second best, and the more grouped method least efficient;

(2) as regards the time required to learn, while there is very little difference, the grouped method is best, on the whole.

Considering all four methods, we find that as regards readings and recitations the mixed method is best. As regards time however, it is not so good as the Method III.

Explanation.—The "mixed method" probably derives one advantage from the fact that there are more readings to start with. However, as it utilizes all three other methods, but Method I. most, and inasmuch as Method I. is most efficient in comparison to Method II. and III., it seems reasonable to think that Method IV. gets its second advantage through involving the interspersed method.

Sex differences.—A glance at the tables (XVI. and XVII.) and graphs IV. and V. shows that the women do much better than the men in:

(1) the number of readings required,

(2) the number of readings and recitations combined, and,

(3) the relative amount of time taken to learn.

Also, the men are far more variable than the women, as shown by the mean variations. In a word, the women surpass the men in memorizing poetry and the men are more variable.

Part III. *R* Used in a Mixed Sense

For comparative purposes, and to test our methods as they would be involved in a practical learning (*e.g.*, by the student), we now used *R* in the mixed sense, that is, an *R* was an attempted reconstruction plus a prompting (visual) when the subject was 'stuck.' The *L* remained the same as previously used. The subject would, *e.g.*, read through according to the method and then try to recall all possible, taking a look to help himself along. Obviously the first *R*'s were largely readings but became progressively more truly *R* as the process continued. Table XIX. gives the results, averages and mean variations.

TABLE XVIII

SHOWING COMBINED RESULTS FOR ALL METHODS FOR 19 MEN AND 17 WOMEN

	Method I.			Method II.			Method III.			Method IV.		
	L's	*LR*'s	Time	*L*'s	*LR*'s	Time	*L*'s	*LR*'s	Time	*L*'s	*LR*'s	Time
Total....	310.35	621.45	504.49	349.32	678.75	582.59	380.40	735.10	449.12	276.27	586.52	470.58
Av.....	8.62	16.98	14.41	9.92	19.39	14.64	10.56	20.41	12.83	8.12	17.25	14.26
M.V.....	2.65	2.45	3.61	3.05	6.09	3.83	3.28	6.38	3.71	2.51	4.94	4.27

TABLE XIX

SHOWING COMPARATIVE RESULTS IN I., II., III., AND IV., FOR 11 MEN

Subject No.	Method I.			Method II.			Method III.			Method IV.		
	L's	*LR*'s.	Time.	*L*'s	*LR*'s.	Time.	*L*'s	*LR*'s.	Time.	*L*'s	*LR*'s.	Time.
41.....	8.70	17.40	17.70	7.30	15.00	13.70	7.00	14.30	13.00	6.30	13.00	11.30
42.....	6.50	13.00	12.25	6.50	13.00	12.75	8.25	16.50	17.00	7.25	14.50	12.75
43.....	6.50	13.00	18.00	8.00	16.00	18.30	7.30	14.60	17.30	6.75	13.50	18.00
44.....	5.00	10.00	9.75	6.70	13.00	12.00	6.75	16.25	12.10	6.25	12.00	12.50
45.....	4.00	8.00	9.10	4.50	8.75	9.90	5.25	9.50	8.80	4.50	8.75	10.20
46.....	7.50	15.00	15.25	8.70	16.70	15.70	7.50	14.75	15.00	7.50	14.50	12.25
47.....	5.00	10.30	9.70	5.50	10.00	8.00	3.00	6.00	8.00	—	—	—
48.....	7.70	15.40	15.00	7.30	14.30	13.75	7.70	15.40	16.00	9.30	18.30	14.30
49.....	4.70	9.40	10.00	6.30	12.30	8.80	6.70	12.00	7.20	5.00	10.00	6.70
50.....	6.30	12.60	13.70	4.00	8.00	10.70	6.00	10.70	11.70	5.30	10.30	10.70
51.....	5.25	10.50	11.90	6.50	12.50	11.50	4.25	8.25	8.75	5.50	9.50	7.00*
Total....	67.15	134.60	142.40	71.30	139.60	135.10	69.70	138.30	134.90	63.65	124.40	115.70
Av.....	6.10	11.23	12.94	6.48	12.68	12.28	6.33	12.57	12.26	6.35	12.43	11.57
M.V.....	1.18	2.45	2.78	1.23	2.16	2.32	1.25	3.07	3.12	1.06	2.32	2.49

The results are not very satisfactory, because the differences are not large. However, of all four methods, Method I., interspersed, is the best as regards the number of readings and readings-recitations combined. As regards time, the order of efficiency is IV., III., II., I.

General Conclusion

Using nonsense material and sense material, the interspersed method of reading recitation seems to be better than methods of grouping. The results are more striking for nonsense material than for sense material.[1]

The question still unsettled is as to how many preliminary readings one should make before introducing the interspersed method of reading recitation. Gates has answered this question probably better than any one else so far, but the optimum preliminary readings will vary according to the length of the material, its difficulty and organization. Certainly introduction of the recitations too early does not favor the interspersed method of reading-recitation.

[1] The experiments reported in this paper were performed in the Psychological Laboratory at the University of Michigan.

RESULTS OF SOME EXPERIMENTS ON AFFECTION, DISTRIBUTION OF ASSOCIATIONS, AND RECALL

BY C. H. GRIFFITTS

University of Michigan

The data presented in this report represents the result of some preliminary spade-work which has for its goal a means of measuring individual differences in cheerfulness. The immediate stimulus for the first experiment was the appearance in 1917 of the report by Baxter, Yamada, and Washburn,[1] which not only suggested a possible means of measuring temperaments, but also served to re-open the question of the relation between affection and recall. Later, the same year, the present writer gave a free, discrete association test three times to the same subjects, who designated as '*P*,' '*U*,' or '*I*' the nature of the association in each case, hoping to obtain information of the following points: (*a*) the relation between affection and recall, (*b*) the relation between scattering or concentration of associations for the three presentations, and recall, (*c*) the relation between this distribution of associations and the kind of affective response, and (*d*), the existence or non-existence of a general affective sensitivity. Nothing more was done until the last school year, when other similar experiments were given, partly to check up on the results of the first experiment, and partly to determine the results of correlating the scores of these later tests with self-estimates in cheerfulness. These experiments will be described later.

Baxter, Yamada, and Washburn, in the article just mentioned, reported the results of an investigation in which they used two lists of stimulus-words, to one of which the subjects were required to get pleasant associations, to the other, unpleasant associations. The ratios of the average reaction

[1] 'Directed Recall of Pleasant and Unpleasant Experiences,' *Amer. J. of Psychol.*, 1917, **28**, 155–157.

times of the two lists furnished an index which showed some correlation with the estimates of friends who judged the subjects in cheerfulness.

In 1919, Morgan, Mull, and Washburn[1] used a different method, which is best described by the instructions to their subjects. "When I pronounce a word to you, observe what idea that word first calls to your mind, and report whether it is a pleasant or unpleasant idea. If it is neither pleasant nor unpleasant, but indifferent, continue thinking until either a pleasant or unpleasant idea is suggested, and report what it is." No times were recorded, the scores being the number of pleasant and unpleasant words suggested. These scores showed some correspondence with the estimates of friends.

The results of these experiments make possible another explanation of most of the data which have been presented in favor of the theory that unpleasant experiences are repressed, a theory of considerable importance in connection with tests of cheerfulness and depression. In general, three sorts of evidence have been advanced in favor of this theory, one of which is represented by the work of Henderson[2] and of Myers,[3] who asked subjects to recall a number of experiences, or of objects of a certain class, and to indicate which were pleasant or were liked, and which were not. The results agree in showing a greater number of pleasant than of unpleasant objects. But these results may be the result of the operation of any one or more of at least three factors: first, the repression of the unpleasant as such; second, a greater number of pleasant than of unpleasant experiences in the previous life of the subject, and third, the mood of the subjects, most of whom are likely to be rather more cheerful than depressed.

Another line of evidence has been secured by testing the

[1] 'An Attempt to Test Moods or Temperaments of Cheerfulness and Depression by Directed Recall of Emotionally Toned Experiences,' *Amer. J. of Psychol.*, 1919, 30, 302–304.

[2] 'Do We Forget the Disagreeable?' *J. of Phil., Psychol., & Sci. Methods*, 1911, 8, 432–438.

[3] 'Affective Factors in Recall,' *J. of Phil., Psychol., & Sci. Methods*, 1915, 12, 85–92.

ability to memorize lists of pleasant and of unpleasant words. Tait[1] found that ordinarily the 'pleasant' lists are more readily memorized than either 'unpleasant' or 'indifferent' lists. In most cases the memorability of the 'unpleasant' lists is higher than that of the 'indifferent' lists. The results, therefore are not regarded as conclusive evidence that the unpleasant is repressed. They might conceivably be due either to the greater number of pleasant experiences, or to the mood of the subject. That these may be factors is suggested also by the fact that for some subjects the 'unpleasant' lists are remembered best.

Tolman,[2] using a similar method, found that an average of 11.5 repetitions were required for the 'pleasant' lists, 12.7 for the 'indifferent,' and 13.5 for the 'unpleasant.' His results differ from Tait's in placing the memorability of the indifferent words above that of the unpleasant words.

Somewhat different is the evidence submitted by Tolman and Johnson,[3] who determined the average association reaction times to '*P*,' '*U*,' and '*I*' words. They found that the times were longer when the '*U*' words were given. Here we may be dealing not so much with a repression of the unpleasant as with a disturbance due to the affective or emotional reaction. The same result might also be expected if we assume that there are more pleasant than unpleasant experiences in the lives of the students who generally act as subjects in these experiments. Tolman and Johnson found their most positive results in the words referring to simple sensory qualities, and it seems to the writer that more of the pleasant than of the unpleasant words used by them would call up experiences which the average college student has actually had. If so, a difference in the reaction times of the two lists would naturally be the result.

Peters[4] used a discrete association method, requiring his

[1] 'The Effect of Psycho-physical Attitudes on Memory,' *Studies in Abnorm. Psychol.*, Series 4, 10–38.

[2] 'Retroactive Inhibition as Affected by Conditions of Learning,' *Psychol. Monog.*, 1917, **25**, No. 107, pp. 22 ff.

[3] 'A Note on Association-time and Feeling,' *Amer. J. of Psychol.*, 1918, **29**, 187–196.

[4] 'Gefühl und Erinnerung,' *Psychol. Arb.* (Kraepelin), 1911, **6**, 197–260.

subjects to respond to the stimulus words by recalling some previous experience, followed by the description of the feeling tone of the original experience as well as of the memory itself. He found more pleasant than unpleasant associations. Perhaps there were more to be recalled.

In all the above experiments,[1] a difficulty arises from the fact that we cannot always control the nature of the affective reaction by our choice of stimulus-words. The result of this, however, would be rather to obscure results than to give spurious results. In the experiments about to be described an effort was made to get around this particular difficulty.

Kate Gordon[2] showed a series of figures to her subjects and later tested their ability to recognize the ones previously shown. She concludes that "Das Wiedererkennen der früher gesehenen Figuren ist gleich gut für die gefälligen, missfälligen und gleichgültigen."

In the light of the above experimental data and of the various possible interpretations in most cases, no definite answer can as yet be given to the question regarding the repression of the unpleasant.

Method

Experiment I.—A list of fifty words, half adjectives and half nouns was prepared, about equally divided between those which might be expected to call up pleasant associations and those which might be expected to call up unpleasant associations. The instructions to the subject were as follows: "I have a list of fifty words, which will be pronounced one at a time at intervals of thirty seconds. When a word is given, you are to wait until it suggests some previous experience when you are to write in the proper space some word connected with the associated idea. You will also record, in the third column, whether the association called up was pleasant, unpleasant or indifferent." The nouns were given first, fol-

[1] No attempt has been made to cover the whole field of literature bearing on this subject: representatives only of the different methods have been selected. Tait gives a rather extensive bibliography.

[2] 'Ueber das Gedächtnis f. affectiv bestimmte Eindrücke,' *Arch. f. d. ges. Psych.*, 1905, 4, 437–459.

lowed immediately by the adjectives. The whole set was given three times, in immediate succession, in different order. Three separate responses were therefore obtained for each word. Then, after a series of group mental tests had been given, requiring about an hour, the subjects were asked to recall as many as possible of the stimulus words used in this test. This was wholly unexpected by the subjects. There were thirty-six subjects, all of whom were beginning students in psychology.

A separate blank had been prepared for each presentation, with a numbering system to facilitate the tabulation of the nature of each of the three responses to each stimulus-word. It may readily be seen that any one word might evoke three '*P*,' '*U*,' or '*I*' responses, two of any one response and one of either of the other two, or one each of the three. There were therefore ten possibilities to consider in tabulating the nature of the affective responses.

It is also evident that the reaction-word (if such it may be called) might be the same in each of the three responses; that there might be but two different reaction-words, one of these being repeated; or that there might be three different associations. Since the distribution of associations might affect the recall of the stimulus word, and because it seemed advisable to determine the relation between this distribution of associations and the nature of the affective responses, three tables were prepared, headed D_1 (the same reaction-word each of the three times); D_2 (two associations, one twice, the other once); and D_3 (three separate associations, each appearing once). Under each of these three main headings, there are ten columns, one for each of the possibilities noted in the preceding paragraph. In tabulation, the results for each subject were kept separate, by using a different row in the tables for each subject. This row was divided by a fine line. If the stimulus word were recalled, the mark in the table indicating the particular combination of affective and associative (ideational) responses was placed above this line; if the stimulus word were not recalled, this mark was placed below the line. This large table, when completed, was to-

talled both by columns and by rows. The totals by rows indicated the number recalled and forgotten by each subject, for each of the three possible distributions of associations. The totals of the columns showed the number remembered and forgotten for each of the possibilities with regard to affective responses; that is, the totals showed the number of marks 'above the line,' and the number of marks 'below the line.' The totals of the affective responses were determined for each subject by a system of points. A mark in the P_3 (each response pleasant) column counted three points for P; a mark in the P_2–U_1 column counted two points for P and one for U, since this meant that two of the three responses were pleasant and the other unpleasant. This was done separately for the marks 'above the line' and 'below the line' (stimulus word recalled, and stimulus word not recalled). Three columns of totals were the result, one each for P, U, and I responses. Three additional columns showed the ratio of the points above the line to the points below the line for each subject, and for the P, U, and I columns respectively. These ratios were used as a measure of the effect on recall of each of the three possible affective responses, although separate results will be given where the P_3, U_3, and I_3 columns only were considered.

It may be objected that the actual distribution of experiences recalled can not be accurately determined in the manner described above, and the justice of the criticism must be admitted. Yet the experimenter believes that, on the whole, the results are approximately correct. The errors which doubtless crept in would not be likely to produce fictitious correlations.

Another criticism may be that no consideration was given in the instructions to the subject to the rather important distinction to be made between the affective tone of the original experience which was recalled and the affective coloring of the recall. Nothing was said about this distinction to the subjects for fear that the net result would be to confuse more than anything else, since most of the subjects were just beginning the study of psychology. It is probable that

in the majority of cases the judgment was based more on the nature of the original experience than upon the feeling tone accompanying the recall. Yet in the majority of cases this affective tone was probably the same, though of less intensity, in the recall as in the original experience. For convenience, in the discussion to follow, I shall speak of the 'affective response' or the 'affective reaction' to the stimulus-word.

The pronounciation of three of the one-syllable words was misunderstood by several of the subjects and were eliminated in the tabulation of the results.

Experiment II.—This experiment was similar to the preceding, differing in that 100 words were used, in that the words were presented visually, and in that the list was given but once. The subjects were later asked to recall the stimulus-words, as in the first experiment. The tabulation of the results was the same except for the simplification arising from the fact that there was but one presentation.

Experiment III.—This was a duplication of Experiment II., save for the fact that a different list of words was used. There were only thirteen subjects in this group.

Experiment IV.—Continuous association test. The subjects were first given a free, continuous association test; then a three-minute test in which they were to write as many as possible of words with pleasant meanings or associations; and another similar period during which 'unpleasant' words were to be written. The number of words written in each of the three parts of the test were tabulated, and the different ratios between the scores computed for each subject.

V. Questionnaire.—The subjects were asked to rate themselves in cheerfulness, as Very Low, Low, Average, High, or Very High. The value of such estimates is obviously limited. Yet this method was used in the hope that any very close correlations would show up to a degree in such results. Most of the subjects were strangers to each other. This was a part only of a more extensive questionnaire, which included self-estimates in twenty-five traits. The rest of this questionnaire was given for other purposes and need not concern us here.

Experiments II. and IV. and the Questionnaire were given to the same group of sixty subjects, all beginning students in psychology. Coefficients of correlation were computed between the results of these tests.

Results

The results will be presented and discussed by topics rather than by experiments, since this will save duplication which would otherwise arise from the similarity of most of the experiments. Those results for which we must look to the first experiment alone are discussed first.

1. The relation between the nature of the affective responses and the distribution of associations (Experiment I.). The score for each subject for the distribution of associations was the total number of different associations called out by the stimulus words. This was obtained by multiplying the number of the $D1$ cases by one, the $D2$ cases by 2, the $D3$ cases by three, and by adding the three products. The total number of points for P, U, and I responses respectively, also the ratio of pleasant to unpleasant responses were then correlated with the scores for distribution of associations with the following results:[1]

Distr. of Associations, and P responses	−.20
Distr. of Associations, and U responses	.53
Distr. of Associations, and I responses	−.17
Distr. of Associations and U/P responses	.39

These figures indicate a fairly close relationship between the number of unpleasant responses and the number of different associations. This is also clearly evident in the data in Table I.

Yet if we turn to Table II., showing the tabulation of the $P3$, $U3$, and $I3$ responses only, we find no evidence of any such relationship. If, in the '$P3$' column, we multiply the number in the $D1$ row by one, the number in the $D2$ row by 2, and the number in the $D3$ row by three, add the three products

[1] No attempt was made to correct for attenuation any of the coefficients presented in this report. The short-cut method proposed by Adams (*Psychol. Bull.*, 1918, **15**, 456–460) was used. This makes use of a formula derived by Thurstone (*Psychol. Bull.*, 1917, **14**, 28–32) from the familiar 'products-moments' formula.

and divide by the total number of $P3$ responses, and then do the same with the $U3$ and $I3$ columns, we get the following values: $P3$, 1.85; $U3$, 1.86; $I3$, 1.73. This means that those words with three pleasant responses called up, on the average, 1.85 different associations, etc., there being no difference in the number of different associations called up by the $P3$ and the $U3$ stimulus-words. However, we should not expect to find the same difference in the results shown in Table II. that we find in the results shown in Table I. where there was a change in the nature of the affective responses to many of the stimulus words. In this connection it should be remembered that the data presented in Table I. include the data shown in Table II.

TABLE I

SUMMARY OF RESULTS OF EXPERIMENT I., BASED ON THE TOTAL NUMBER OF POINTS IN EACH CASE

	P	*U*	*I*	Total
*D*1, Number	511	371	489	1,371
% recalled	64.3	59.8	50.5	58.2
*D*2, Number	735	638	655	2,028
% recalled	62.0	58.2	56.8	58.8
*D*3, Number	593	506	578	1,677
% recalled	64.2	64.2	62.6	63.6
Total, Number	1,839	1,515	1,722	5,076
% recalled	63.3	60.5	56.6	60.2

TABLE II

SUMMARY OF RESULTS OF EXPERIMENT I., BASED ON TOTALS OF POINTS, BUT LIMITED TO $P3$, $U3$, AND $I3$ CASES

	*P*3	*U*3	*I*3	Total
*D*1, Number	142	104	126	372
% recalled	66.2	60.5	50.0	59.2
*D*2, Number	146	121	111	378
% recalled	62.4	62.8	52.3	59.5
*D*3, Number	86	63	49	198
% recalled	71.0	65.0	59.2	66.2
Total, Number	374	288	286	948
% recalled	65.7	62.5	52.4	60.8

2. The distribution of associations and the recall of the stimulus-word (Experiment I.). Direct evidence of a positive correlation may be found in Tables I. and II. The '*I*' column in Table I. and the '*I*3' column in Table II. contain

data which are not affected by the nature of the affective response to the stimulus-words, and in these columns there seems to be clear evidence of a direct positive relationship between the number of different associations called out by a stimulus-word and the probability that that word will later be recalled.

In addition, the results obtained by the use of adjectives and those obtained by using nouns were tabulated separately, to show the per cent. recalled for each of the three distributions ($D1$, $D2$, and $D3$) with the following results:

Adjectives.....$D1$, 58.6; $D2$, 64.4; $D3$, 72.5.
Nouns.....$D1$, 56.7; $D2$, 53.4; $D3$, 56.5.

The coefficient of correlation between the number of different associations and the total number of all stimulus words recalled is only .18. The method of partial correlation, taking the affective response into account raises this to .38. Yet the small size of these coefficients is not to be regarded as evidence against the existence of a real correlation.

3. Distribution of affective responses. On this point the results are in accord with those reported by other investigators. In Experiments I., II., and III. (Tables I., II., and III.), the results agree in showing a greater proportion

TABLE III

DISTRIBUTION OF AFFECTIVE RESPONSES AND THE PER CENT. OF STIMULUS-WORDS RECALLED IN EACH CASE. BASED ON 'POINTS' IN THE ORIGINAL TABULATION OF DATA

Response	Experiment II		Experiment III	
	Av. Number	Per Cent. Recalled	Av. Number	Per Cent., Recalled
P.....	41.1	30.0	27.3	25.2
U.....	31.0	24.6	17.6	23.1
I.....	27.8	27.4	10.1	23.0
U/*P*.....	78.8	84.6	57.5	71.0

of pleasant than of unpleasant associations. Since different sets of words were used in each of the three experiments I believe it improbable that the results are due to an unfortunate choice of words. Yet, as has been pointed out earlier in this report, such results offer no evidence that unpleasant experiences are repressed.

The results of Experiment IV., shown in Table IV., show that more 'pleasant' than 'unpleasant' words may be written in three minutes.

TABLE IV

(EXPERIMENT IV)

Average Number of Words per Subject Written in Three Minutes

	Free	*P*	*U*	*U*/*P*
Mean.....	62.8	36.5	31.7	.88
S.D.....	14.7	12.6	11.8	.23

4. The effect of the affective response upon the recall of the stimulus-word. The most direct evidence on this point is to be found in Tables I., II., and III., showing the per cent. recalled under different conditions. In every case a greater per cent. of *P* than of *U* stimulus-words were recalled, and, generally, a greater per cent. of *U* than of *I* words. However, these results are not necessarily to be regarded as evidence that unpleasant experiences and the things associated with unpleasant experiences are repressed. They indicate rather that an affective reaction of any kind is favorable to the recall of the stimulus-word. The fact that a greater per cent. of words calling up pleasant associations than of those calling up unpleasant associations were recalled may indicate either that a pleasant reaction is more favorable to recall than an unpleasant reaction, or else, that the responses designated as pleasant were, on the whole, stronger than those designated as unpleasant.

For some reason the nature of the affective response seems to be more effective upon recall when adjectives are used as stimulus-words than when nouns are used (Experiment I.). The per cents. of *P*, *U*, and *I* stimulus-words respectively, which were recalled in each case were as follows: For adjectives, 74, 64, and 54; for nouns, 55, 56, 54.

Coefficients of correlation were computed between the numbers of *P*, *U*, and *I* responses, respectively, and the number (not per cent.) of *P*, *U*, and *I* stimulus-words recalled. The resulting coefficients were; *P*, .68; *U*, .83; *I*, .86. These would be quite as large if not larger in the absence of any

effect upon the recall of the stimulus-word by the affective reaction.

I have assumed that ordinarily there was some affective reaction accompanying those associations designated by the subjects as '*P*' or '*U*.' Physiological reactions observed during similar experiments make such an assumption reasonably safe. If one should make the opposite assumption, the results would indicate at least that those words which call up experiences which were affectively toned are more likely to be recalled than other words.

5. Another question of considerable practical importance in the devising of tests of temperament has to do with the relation of the percentage of *P*, *U*, or *I* responses and the percentages of the stimulus-words recalled for each of the three kinds of response. For example, will the factors which cause one person to have more pleasant reactions than the average of the group, also cause him to rank high in the group with regard to the ratio of the percentage of *P* stimulus-

TABLE V[1]

INTRA-TEST CORRELATIONS

Association	Per Cent. Recalled	Experiments		
		I	II	III
P	*P*	.06	.08	−.10
U	*U*	−.19	.14	.30
I	*I*	.30	−.13	.00
U/P	*U/P*	−.16	−.18	−.32
U/I	*U/I*	.39	.26	.20
P/I	*P/I*	.07	.10	.13

words recalled to the percentage of *U* stimulus-words recalled? If the kind of affective reaction is affected by the mood of the subject (association by congruity of emotional tone), or by a greater number of pleasant or of unpleasant experiences, or by a repression of the unpleasant, it would seem reasonable

[1] The method of tabulation 'by points' has been explained under 'Method.' In tabulation the results from each stimulus-word for each subject, the mark in the table representing the particular combination of affective responses and 'distribution of associations' was placed above a line if the word were later recalled, below the line if it were not recalled. The 'Per Cent. Recalled' represents the ratio of the number above the line to the whole number.

that the same factors would affect the memorability of the stimulus-words for the different kinds of affective response. Yet the coefficients of correlation shown in Table V. offer little positive evidence either way.

When there were a very few affective reactions of a certain kind, ordinarily a greater percentage of these than of the other stimulus-words were recalled. This may be because they were something of a novelty. If so, this novelty would be a factor operating against any effect of the mood of the subject in determining recall. It may also be true that there is something of a negative reaction away from those words calling up unpleasant memories. But if so why should there not be more evidence of a repression of the stimulus-words evoking unpleasant responses? The answer may be that this tendency, if it exists, is overcome by the operation of other factors, particularly the favorable effect upon recall of any affective response. The fact that the recall value of the *P* stimulus-words is higher than that of the *U* stimulus-words, and that it seems to be higher for both than for the *I* stimulus-words would fit in with this explanation. It remains for further work to isolate these factors more completely and determine which are operative and when.

6. Is there a general affective sensitivity? Is a high threshold for one kind of affective reaction associated with a high or low threshold for the other kind of reaction? Some writers seem to assume an affirmative answer, others a negative. While the question is fundamental, it is one which is rather difficult to answer.

If, in association experiments, there were always some affective reaction, it is evident that the coefficient of correlation of the number of pleasant responses with the number of unpleasant responses would be minus one. Yet if we could measure the strength of the reactions we might find a fairly large positive correlation. In any event, the negative correlation between the numbers of the different reactions could not be regarded as evidence against the existence of a general affective sensitivity. In my experiments, however, there were many cases regarded as 'indifferent' by the subjects.

This makes it possible to obtain some information on this point by correlating the scores (frequency) for the different affective reactions with each other, and each with the number of '*I*' responses. This was done with the data from Experiments I. and II. It should be remembered that there were twice as many subjects and twice as many words used in the second experiment, and that each word was given but once instead of three times as in the first.

TABLE VI

INTRA-TEST CORRELATIONS

	A. Experiment I				*B*. Experiment II		
	P	*U*	*I*		*P*	*U*	*I*
P.....	—	−.02	−.66	*P*.....	—	.26	−.75
U.....		—	−.75	*U*.....		—	−.78
I.....			—	*I*.....			—

C. Experiment I			
U/I and *P/I*.....	.76	*U/P* and *I*.....	−.09
U/I and *U/P*.....	.49	*U/I* and *P*.....	.39
P/I and *P/U*.....	.14	*P/I* and *U*.....	.45

In considering these correlations, shown in Table VI., it must also be remembered that we are still moving in a closed circle, so to speak, since the sum of the reactions is always the same. This being true, an increase in the number of *P* responses must be associated with a decrease in the sum of the *U* and *I* responses. If, therefore, the number of *U* responses should remain the same under such conditions, it would indicate an actual lowering of the threshold for these reactions. Therefore, even a zero correlation between the frequencies of *P* and *U* responses, with a large negative correlation between the *P* and *I*, and between the *U* and *I* responses, would point to an actual positive correlation between the positions of the two thresholds. The writer therefore believes that the coefficients given in Parts *A* and *B* of Table VI. may be regarded as evidence of the existence of a general affective sensitivity.

The coefficients shown in Part *C* probably furnish further evidence along the same line. One reason for showing these

coefficients is their significance with regard to methods of scoring the results of such experiments. Tolman and Johnson used the ratios P/I and U/I in working up their results. But since the results may be affected by the mood of the subject (association by congruity of emotional tone), by a general affective sensitivity, and possibly by a repression of the unpleasant, care must be taken in the selection of the ratios to be used. To bring out the relative positions of the P and U thresholds the ratios P/U or U/P will be found valuable. The use of any such ratios will be made safer by reducing them to logarithms.

I have been speaking of two thresholds, one for P and one for U reactions. It may be that there is but one.

I had hoped that further information might be secured by correlating the results of the three parts of Experiment IV. (continuous association tests, three minutes for each part). The following coefficients were obtained: Between the Free and the 'Pleasant' tests, .59; between the Free and the 'Unpleasant' tests, .42; between the 'Pleasant' and the 'Unpleasant' tests, .61. Since the last one mentioned is so nearly the same as the other two the only safe conclusion is merely that some constant factor or group of factors was operating.

Other evidence of the existence of a general affective sensitivity is to be found in the data presented in Table VII., showing the distribution of affective responses for each of the three presentations of the stimulus-words in Experiment I.

On the whole, the writer believes that the data just discussed indicates the existence of a general affective sensitivity. In other words, there seems to be a positive and not a negative relation between the positions of the two thresholds, the one for pleasant reactions, the other for unpleasant reactions. Common experience tells us that the person who is most likely to show pleasure in one situation is most likely to show displeasure in another situation; the stolid, phlegmatic person is not so likely to show any evidence of an affective reaction in either. Yet, on the other hand, the mood of the individual does play a part, and the same thing may at one time cause pleasure and at another displeasure. In one mood

one threshold seems to be much higher than the other; in the opposite mood, the other threshold seems to be the higher. In abnormal cases the apparent opposition between mania and melancholia would seem to point to a negative relation.

7. The effect of repetition upon affective sensitivity.[1] The number of *P*, *U*, and *I* responses for each of the three presentations in Experiment I. is shown in Part *A* of Table VII.

Part *B* of this same table shows the results of treating the data in another way. For each subject the order-of-merit method was used in ranking each of the three presentations with regard to the number, first of *P*, then of *U*, and finally of *I*, responses. These were then totalled, with the results shown in the table (the larger the number the lower the score).

TABLE VII

A. DISTRIBUTION OF AFFECTIVE RESPONSES
(1,692 Cases, Each Presentation)

	First Presentation	Second Presentation	Third Presentation
	%	%	%
P.....	35.9	37.6	36.4
U.....	29.4	29.9	30.1
I.....	35.4	33.0	33.4
B. SAME; ORDER-OF-MERIT METHOD			
P.....	79.0	70.0	67.0
U.....	75.5	75.0	65.5
I.....	63.5	80.0	72.5

These results are not very conclusive, but suggest a greater affective sensitivity for the second presentation, while that for the third is less than that for the second but more than that for the first. It may be noticed that the number of *P* responses is greater the second presentation, but that there is a steady increase in the number of *U* associations. The experimenter felt that some of the subjects were becoming 'bored' toward the close of the experiment.

8. As stated earlier in the paper, the purpose in giving Experiment IV. was to determine the correlation, if any, between the results of this test and the results of Experiment II.

[1] Professor H. F. Adams first suggested to the writer the desirability of re-tabulating the data to obtain information on this point, because of its importance in the psychology of advertising.

The value of this correlation is affected by the fact that it was not possible to give the two experiments the same day but on days a week apart. The following coefficients were obtained between the results of the two tests: for *P*, .18; for *U*, .09, for *U*/*P*, .30. The last may be of some significance. These results are not very encouraging, although the writer intends to include this in a new series of experiments along the same line the coming school year.

9. Questionnaire. Correlations were computed between the self-estimates in cheerfulness and the different scores in Experiments II. and IV., the only experiments in which the subjects were asked to rate themselves. The correlations were generally positive with the scores based on the number of pleasant associations or on the number of 'pleasant' words written in three minutes, and ordinarily negative with the scores based on the number of unpleasant associations or on the number of 'unpleasant' words written in three minutes. However, the only correlation large enough to be of much significance was one of .47, P.E., .06, between self-estimates in cheerfulness and the ratio of pleasant to unpleasant responses in the discrete association test (Experiment II.).

Summary of Results

1. When the same stimulus-words are presented three times in a free, discrete association test there is a correlation of .41 between the number of unpleasant associations and the distribution of associations (number of different associations).

2. There seems to be a correlation between the number of different associations, when the stimulus words are given three times each, and the probability that the stimulus-word will later be recalled. This correlation is fairly large when adjectives are used as stimulus-words, but is practically zero when nouns are used.

3. In every experiment there were more pleasant than unpleasant associations.

4. There is but little, if any, evidence that stimulus-words calling up unpleasant associations are repressed. To the contrary, the evidence indicates that an affective reaction of

any kind is favorable to the recall of the stimulus. A greater per cent. of the words arousing pleasant associations are recalled than of those arousing unpleasant associations. The memorability is least for those with indifferent associations.

5. The coefficients show very little, if any correlation between the numbers of *P*, *U*, and *I* responses and the percentages of the corresponding stimulus-words which were recalled. A subject with more pleasant than unpleasant responses may or may not recall a greater per cent. of the stimulus-words for the pleasant associations than of the stimulus-words with unpleasant associations.

6. There is evidence of the existence of a general affective sensitivity. In other words, that there is a positive and not a negative correlation between the threshold for pleasant reactions and the threshold for unpleasant reactions.

7. In the first experiment the stimulus-words were presented three times. The affective sensitivity, in so far as it is measured by the sum of the pleasant and unpleasant associations, was greatest for the second presentation, and was higher for the third than for the first.

8. Three minutes were devoted to the writing of as many 'pleasant' words (names of pleasant objects, etc.) as possible, and another three minutes to the writing of 'unpleasant' words. There is some correlation, .31, P.E., .06, between the ratio of the number of words in the 'pleasant' list to the number in the 'unpleasant' list, and the ratio of pleasant to unpleasant associations in a free discrete association test taken by the same group of subjects.

9. The ratio of pleasant to unpleasant associations in a free discrete association test showed a correlation of .47, P.E., .06 with self-estimates of cheerfulness.

TWO NEW TIME CONTROL INSTRUMENTS

BY DONALD A. LAIRD

University of Iowa

The two instruments described below were designed to meet the requirements of a research problem which were not satisfied by any existing forms of time control apparatus. Inasmuch as accessory apparatus for measuring out and controlling time intervals is of a considerable utility in a psychological laboratory in regulating the duration and sequence of stimuli and operations, in actuating certain types of exposure apparatus, in operating a series of instruments as is necessary in the finer measurements of reaction times and in numerous other well-arranged measurements, and in checking the accuracy of other instruments; and since these to be described possess several unique and desirable features over other forms it is well that they be placed at the disposal of laboratory psychologists in general.

I. An Acceleration Time Control Apparatus

Time control apparatus based upon the action of gravity on falling bodies has been in use almost since the beginning of laboratory psychology. In the present device, however, the physical phenomena of acceleration is utilized. The essential construction of the apparatus consists of an inclined U-shaped groove upon which a steel sphere rolls. The inner edges of this groove, which form the track for the sphere, are constructed of alternating strips of insulating and conducting material placed in such a manner that the sphere in its passage along the inclined track completes the circuit through the conducting strips arranged opposite each other.

Knowing the angle of inclination of the groove (ϕ), the distance the sphere travels before coming into contact with the conductors (d^1), the distance the ball travels over the conducting strips (d^2), the radius of the sphere (r), the distance

separating the edges of the strips (s), and the acceleration of gravity (g) it is possible to compute the time required for the passage of the sphere over the conductors by the following formula:[1]

$$T = \sqrt{\frac{1}{g \sin \phi}\left[2 + \frac{4}{5\left(1 - \left[\frac{s}{2r}\right]^2\right)}\right]}\left(\sqrt{d^1 + d^2} - \sqrt{d^1}\right).$$

Theoretically the time consumed by the ball in passing over the conductors could be varied in several ways. The distance over which the ball passes before coming into contact with the conductors could be varied, the actual length of the conductors altered, the angle of inclination increased or decreased, the size of the sphere varied, or the distance separating the edges of the groove changed. In practical experience it was found that the instrument operated with the greatest reliability, *i.e.*, made the most secure contacts, when the inclination of the groove was about 3 degrees. Some of the other possible sources of alteration of the time intervals being measured out are obviously out of the question with a firmly constructed instrument, as, for example, changes in the distance separating the strips. Theoretically, also, if a V-shaped groove, the sides of which formed a constant and known angle, were used the acceleration would be independent of the size of the sphere; but again, as a matter of experience, it was soon found that a much better contact and a more uniform track was secured when the strips were placed on edge.

The point at which the sphere is started on its downward journey is kept constant and the initial velocity kept at zero by means of a release attachment, consisting of an electromagnet arranged so that it can be securely fastened at any point along the track. To start the sphere it is only necessary to break the circuit through the magnet. The best results were obtained when a laminated core was used and a thin mica sheet attached to the end with which the ball comes into contact. This precaution avoids the undesirable in-

[1] This formula was arrived at in collaboration with Mr. Victor Hoersch, of the Department of Physics of the University of Oklahoma.

fluence which any appreciable residual magnetism might have.[1] Certain other precautions should also be noted:

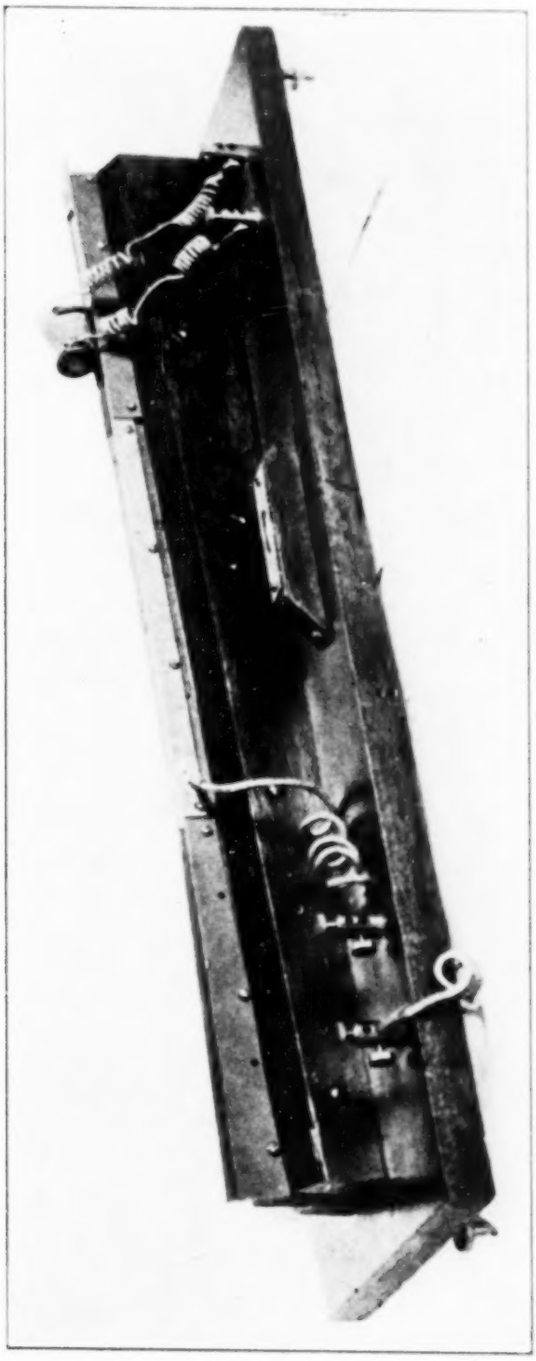


FIG. 1. Acceleration apparatus.

(1) The current must not vary, otherwise the arcing at the bottom of the conductors will produce variations in the

[1] This residual lag could be entirely eliminated if a simple, coreless, magnetic coil were used.

contact length of the strips, (2) with the size and material of the sphere used in the present arrangement a secure contact is not obtained when less than three dry cells are used, (3) the track itself must be kept free from dust, grease, and all foreign matter.

The original instrument as illustrated in Fig. 1 is made of wood with the inclined track of brass and fiber and at an angle of 3 degrees with the base. A one-inch steel sphere is used and the inner edges of the track separated by 12.0 mm. The base is adjustable by means of the three leveling screws and constancy of position is assured by a small spirit level attached to the upper surface. No attempt was made to provide for an adjustment in the inclination of the track as preliminaries had demonstrated the limits of the arrangement. The instrument was planned so that one second would be consumed while the sphere is traveling over the conductors and regardless of the inaccuracies attendant upon amateurish construction it consistently measured intervals of 1.01 second. With a slight adjustment of the position of the release magnet this interval was quickly changed to 1.00 second. The instrument was checked only to the hundredth but it successfully measured intervals between 0.5 and 1.0 seconds accurately within the limits of the checking set up.

The merits of this instrument are:

1. A wide range of times in comparison with other forms of time control devices.
2. A greater flexibility. This instrument is not limited to a make and a break but a variety of electrical combinations is possible in sequence.
3. Accuracy to at least the hundredth.
4. Few adjustments and none to get out of order.
5. Certainty of operation within the limits mentioned.[1]
6. Cheapness and ease of construction. The model illustrated was constructed in six hours. The materials are available around even a modestly equipped laboratory.

[1] Increasing the mass of the rolling body would increase the working limits.

II. A Disc Commutator Time Control Instrument

This instrument utilizes the principle of the staggered commutator combined with the magnetic friction clutch as applied by Dunlap. The chief claim to originality is in the

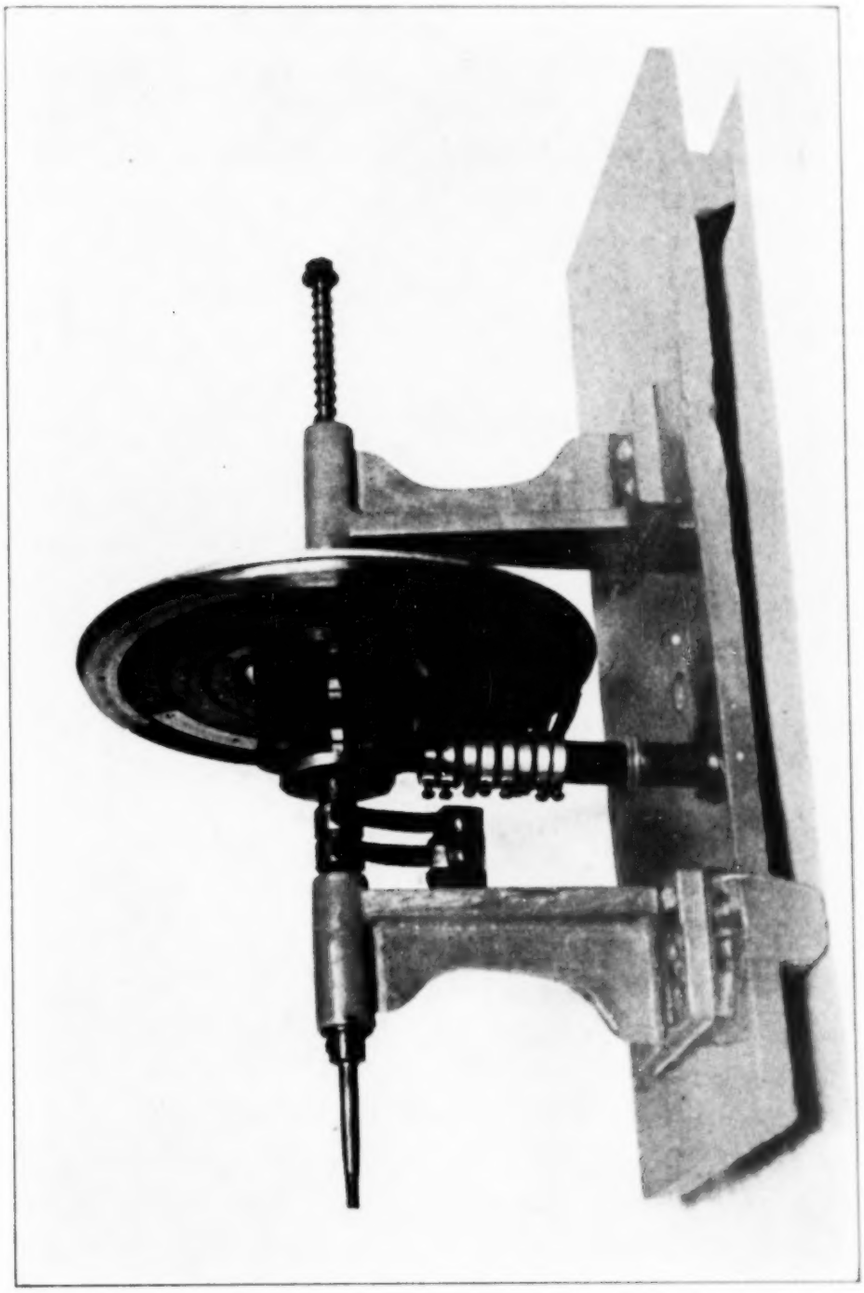


Fig. 2. Disc commutator apparatus.

use of a disc rather than a drum for the commutator and in the matter of the electrical system of the clutch control. The use of a disc makes possible a more compact and a lighter

instrument requiring less power and attention in addition to simplifying the calculations necessary for adjusting the collector sectors for varying time intervals. By the system of electrical connections the instrument can by synchronized with the observer or experimenter, it not being necessary to get the human element into step with a steadily revolving drum or constantly swinging pendulum. This feature is especially desirable in all work with children. The electrical control is such that the apparatus requires no attention and the noise incident to its operation can be removed, with the instrument, into a distant room and all auditory disturbance will be removed from the vicinity of experimentation.

Two shafts placed with their centers opposite each other are mounted on a base (Fig. 2). End thrust is eliminated from the drive shaft by shoulders placed each side of the bushings. One end of the drive shaft is fitted to carry the follower of a set of 45-degree spiral gears which meshes with the driver on the motor shaft. The other end of the drive shaft carries two small electromagnets.[1] A small double ring commutator on the drive shaft carries the current from the brushes to the magnets. The disc shaft is held in place by a shoulder on the end nearest the clutch magnets and by a light compression spring on the other end. The tension of this spring is adjustable by the nut at the end of the shaft. The disc commutator is normally about 1/32 inch from the magnets, so that upon the energization the disc shaft springs laterally 1/32 inch and is firmly gripped by the magnets before beginning to revolve with them. When released by the clutch the tension of the spring brings the disc back into normal position and stops it almost instantly. The two outer overlapping conductors on the disc complete the circuit for the clutch magnets through the two lower pair of brushes. All the pairs of brushes are of four receding leaves of spring brush brass and are always in firm contact with the commutator. The brushes are fully adjustable so that the take off is on a line with the radius of the disc perpendicular to the

[1] In order that eddy currents may be avoided both shafts are supported by aluminum castings.

base. Compensation for wear is provided for in the case of the time brushes (two upper pair) by having the end bent at an angle of 90 degrees.

The disc itself is made of aluminum and is attached to the disc shaft by means of machine screws. Being constructed of a light material it gains less momentum than if it were heavier and consequently less tension is required to bring the disc to a full stop than would otherwise be the case. In

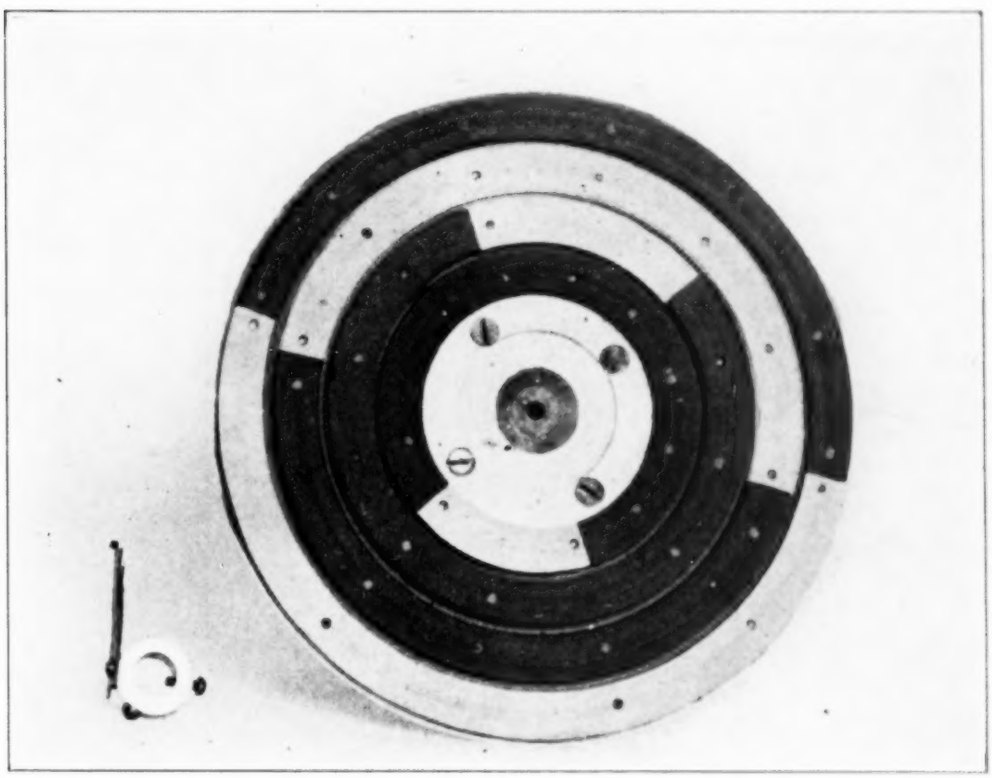


FIG. 3. Face of disc commutator; brush showing holder and adjusting screws.

other terms, less power is used and a smoother operation is attained. The collector sectors are insulated from the disc proper by a sheet of fiber on the face of the disc and fiber washers and bushings are used on the attaching screws.

As arranged in the present instrument the shafts make one complete revolution in 3.6 seconds. At this rate one degree passes the brushes in 0.01 second. Since the time collectors are 70 degrees, seven tenths of a second is measured out in each phase or half revolution.

A wide range of flexibility and a convenient method of change is provided by having holes drilled and tapped in the disc in a symmetrical arrangement on the line of each con-

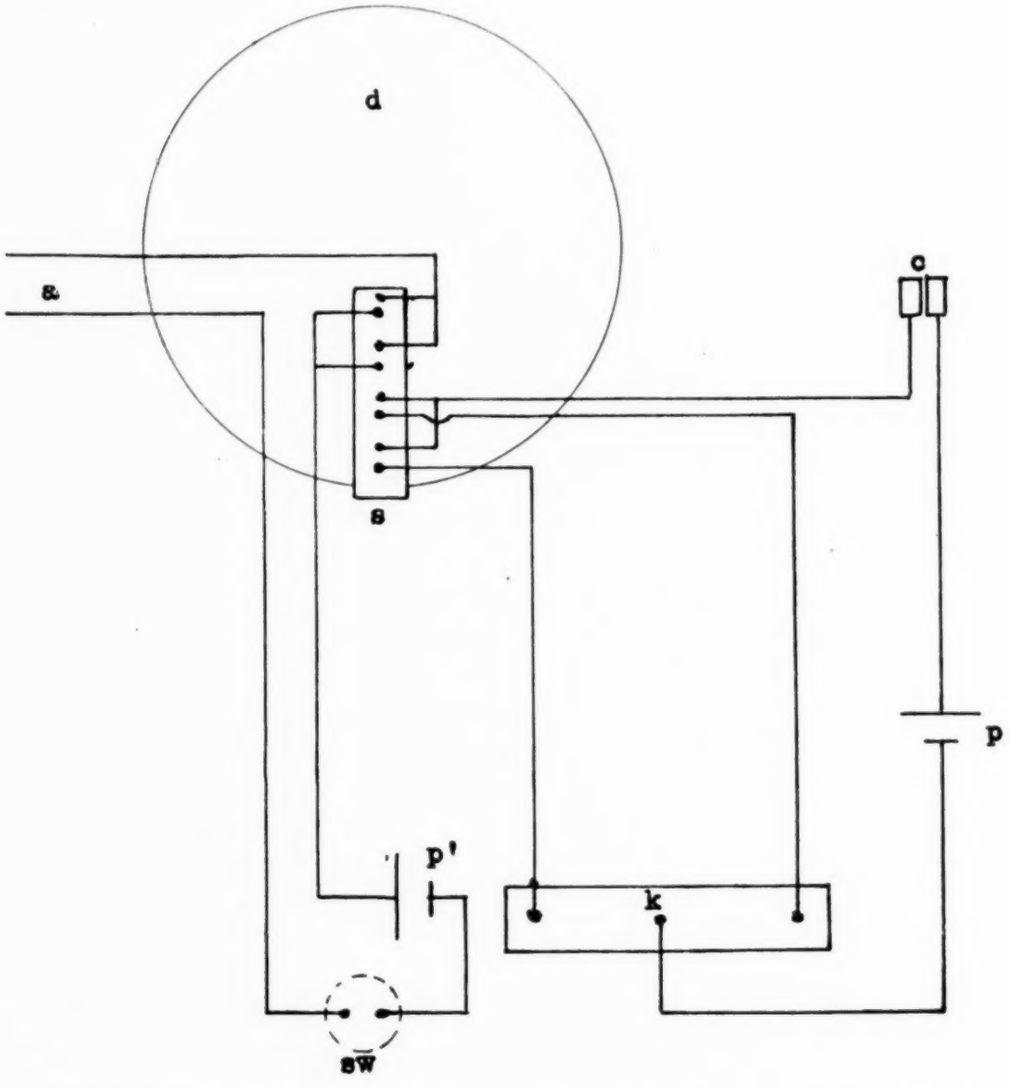


FIG. 4. Diagram of the electrical connections of the disc commutator apparatus. *p* is the source of the primary circuit which passes through the electromagnets by means of the two lower pair of brushes on the brush support, *s*, and the magnet commutator, *c*. *k* is the foot key which shifts the current from the lower to the next above pair of brushes to start the exposure between the phases of revolution of the disc commutator, *d*. *p'* the source of the secondary circuit which passes to the timed apparatus, *a*, through the two inner pair of brushes. *sw*, snap switch for breaking secondary circuit when the control apparatus is not in use.

centric ring. Supplementary sets of collector and insulating rings are threaded all ready to be cut and matched and can

be fastened at any point and for any length on the disc, provided that they are on a line with the circle of openings corresponding to their curvature.

By means of a double throw switch the current can be made to pass through either of the primary (overlapping) collector sectors, as shown in Fig. 4. When the switch is once thrown the disc begins its revolution and continues until its first phase is completed when the circuit is broken by the one primary collector passing the brushes. When the switch is again thrown the other primary conductor is brought into the circuit and the disc again revolves for another phase, completing one revolution. Due to the overlapping of the primary sectors a start is always assured. By this arrangement any initial lag has no effect upon the interval being measured, for the disc travels several degrees before the timing collectors come into contact with their brushes.

Any desired range of speed intervals can be produced by altering the speed of the drive shaft or by changing the sector degrees of the collectors, or by a combination of both. A series of intervals can be produced by providing a series of conducting strips in line with one pair of brushes if it is desired to use the same line, or more than one collector circle can be used.

As a source of power for the motor a 1/6 h.p. motor of the constant speed, repulsion-induction type is recommended. A synchronous motor of the same power could also be used, but the portability and sureness of operation of the former in addition to its reliability and simplicity of operation lead me to prefer the former. No accessory apparatus is needed for the operation of a constant speed motor and alternating current is always available. When operated on alternating current the synchronous motor reacts to the fluctuations in pulsation rate which are constantly occurring and if the motor is delicately adjusted a slight decrease in potential will cause the motor to stop and 'sing.' The higher types of constant speed motors on the other hand, do not react to fluctuations of less than 10 per cent. and very rarely is the electrical service such that this limit is approached and on ordinary

city current a repulsion-induction motor of good construction will operate with a maximum error of only 1/4 of one per cent., even when influenced by the peak load. On an interval of one second this would lead to a possible maximum error of 0.0025 second. It is well known that unless a thermostat is used with a synchronous motor when operated in tandem a gradual but persistent decrease occurs in the rate of rotation.

This disc commutator instrument is more difficult to construct but on the whole is more desirable than the acceleration apparatus for laboratories that can afford the more expensive construction. Its chief points of superiority over the acceleration apparatus and other forms of apparatus for measuring out time intervals are:

1. A range of intervals with a lower limit of about 0.01 second and practically no upper limit.

2. Any combination of connections desired is possible within the time circuits either simultaneously or in sequence.

3. Accuracy. The instrument is as accurate as the source of power.

4. Operation is certain with any currents within the range of laboratory use.

5. Freedom from distraction in the sphere of experimentation.

6. Simplicity of operation. There is nothing to do but throw a switch or press a key when the interval is desired.

7. The intervals can be given at any moment desired and in rapid succession.

The designer of the instrument has the patterns from which the original was constructed and will coöperate with other laboratories in the construction of similar instruments or of adaptations.

PSYCHOLOGICAL REVIEW PUBLICATIONS

Original contributions and discussions intended for the Psychological Review should be addressed to

Professor Howard C. Warren, Editor PSYCHOLOGICAL REVIEW,
Princeton, New Jersey.

Original contributions and discussions intended for the Journal of Experimental Psychology should be addressed to

Dr. John B. Watson, Editor,
244 Madison Avenue, New York.

Contributions intended for the Psychological Monographs should be addressed to

Professor James R. Angell, Editor PSYCHOLOGICAL MONOGRAPHS,
522 Fifth Avenue, New York

Reviews of books and articles intended for the Psychological Bulletin, announcements and notes of current interest, and *books offered for review* should be sent to

Professor S. I. Franz, Editor PSYCHOLOGICAL BULLETIN,
St. Elizabeth's, Washington, D. C

Titles and reprints intended for the Psychological Index should be sent to

Professor Madison Bentley, Editor PSYCHOLOGICAL INDEX,
University of Illinois, Urbana, Ill.

All business communications should be addressed to

Psychological Review Company
41 N. Queen Street, Lancaster, Pa., and Princeton, New Jersey.

PHILOSOPHICAL MONOGRAPHS

VOL. I

1. **Æsthetic Experience: Its Nature and Function in Epistemology.** W. D. FURRY. Pp. xvi+160. $2.60. 2. **The Philosophy of John Norris of Bemerton.** FLORA ISABEL MACKINNON. Pp. 18+105. $1.00. 3. **Freedom and Purpose. An Interpretation of the Psychology of Spinoza.** JAMES H. DUNHAM. Pp. iv+126. $1.25.

LIBRARY OF GENETIC SCIENCE AND PHILOSOPHY

Vol. 1. The Moral Life. A Study in Genetic Ethics. By ARTHUR ERNEST DAVIES, Ph.D., Professor of Philosophy in Ohio State University. xiv+188 pages. 8vo. Cloth. Price, $1.50 net. Postage, 10 cents.

Vol. 2. Darwin and the Humanities. By JAMES MARK BALDWIN. viii+118 pages. 8vo. Cloth. Price, $1.50 net. Postage, 8 cents.

Published by

PSYCHOLOGICAL REVIEW COMPANY

Directory of American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.: Morrill Hall.
Subscription [illegible] annually. Edited by E. B. Titchener.

Pedagogical Seminary—Worcester, Mass.: Pedagogical Chamber.
[illegible] annually. Edited by G. Stanley Hall.
Child study and educational psychology. Founded 1891.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company.
Subscription [illegible] 480 pages annually.
Bibliographical. Founded 1894. Edited by Howard C. Warren.

Psychological Index—Princeton, N. J.: Psychological Review Company.
Subscription [illegible] 300 pages annually. Psychological literature.
Bibliographical. Edited by Shepherd I. Franz.

Psychological Monographs—Princeton, N. J.: Psychological Review Company.
Subscription [illegible] 400 pp. Founded 1895. Edited by James R. Angell
Published [illegible] each issue one or more researches.

Psychological Review—Princeton, N. J.: Psychological Review Company.
Subscription [illegible] 500 pp. Founded 1894. Edited by Madison Bentley
An annual bibliography of psychological literature.

Journal of Philosophy—New York: Sub-Station 84.
Subscription [illegible] 416 pages per volume. Founded 1904.
Bi-weekly. Edited by F. J. E. Woodbridge and Wendell T. Bush.

Archives of Psychology—Sub-station 84, N. Y.: Archives of Psychology.
Subscription [illegible] 150 pp. Founded 1906. Ed. by R. S. Woodworth.
Published [illegible] each number a single experimental study

Journal of Abnormal Psychology—Boston: Richard G. Badger.
Subscription [illegible] 400 pages annually. Edited by Morton Prince.
Bi-monthly. [illegible] Entire field of abnormal psychology.

Psychological Clinic—Philadelphia: Psychological Clinic Press.
Subscription [illegible] 300 pages. Ed. by Lightner Witmer. Founded 1907.
Witmer (8 numbers). Orthogenics, psychology, hygiene.

Training School—Vineland, N. J.: The Training School.
Subscription [illegible] 80 pp. 4to. Ed. by E. R. Johnstone. Founded 1904
Monthly (10 numbers). Psychology and training of defectives.

Journal of Religious Psychology—Baltimore: Warwick & York.
Subscription [illegible] 300 pages annually. Founded 1910.
Managing Editor, J. Cadeton Bell.
(Occasional Monographs. Edited by Guy M. Whipple.
Subscription [illegible] varying prices. Issued [illegible].)

The Behaviorist—Cambridge, Mass.: Emerson Hall.
Subscription [illegible] 400 pages per volume. Edited by John B. Watson.
Bi-monthly [illegible] each number a single research.

Psychological Clinic—Washington, D. C.: 1835 16th St., N. W.
Subscription [illegible]
Bi-monthly [illegible] Edited by W. A. White and S. E. Jelliffe.

Journal of Animal Behavior—Lancaster, Pa.: [illegible] J.
Subscription [illegible] 480 pages annually. Beginning
Bi-monthly [illegible] Edited by John B. Watson.

Journal of Educational Psychology—Warwick & York, Inc., Florence Columbia.
Subscription [illegible] annually. Founded 1910.
Managing Editor, [illegible] Colvin.

Journal of Experimental Psychology—Baltimore: Williams & Wilkins Co.
Subscription [illegible] 500 pages annually.
Bi-monthly [illegible] Edited by J. A. Fulton. Founded